Notes on Stochastic Relations

Ernst-Erich Doberkat
Chair for Software-Technology

Department of Computer Science
University of Dortmund
doberkat@acm.org

January 12, 2006
## Contents

1 Introduction 5

2 Categorial and Probabilistic Aspects of Stochastic Relations 15
   2.1 The Manes Monad ........................................... 17
   2.2 The Giry Monad ............................................ 19
   2.3 Case Study: Architectural Modelling Through Monads ............ 27
   2.4 Case Study: Probabilistic Semantics of a Simple Language .......... 55
   2.5 Bibliographic Notes ....................................... 67

3 Eilenberg-Moore Algebras for Stochastic Relations 71
   3.1 Algebras for a Monad ....................................... 72
   3.2 Characterization Through Equivalence Relations .................... 73
   3.3 Positive Convex Structures ................................ 81
   3.4 Algebras Through Positive Convex Structures .................... 82
   3.5 Examples .................................................. 85
   3.6 The Left Adjoint ........................................... 89
   3.7 Case Study: Derandomization ................................ 92
   3.8 Bibliographic Notes ....................................... 94

4 The Existence of Semi-Pullbacks 95
   4.1 A Road Map ................................................ 96
   4.2 Extending Semi-Pullbacks of Measures .......................... 100
   4.3 The Existence of Semi-Pullbacks ............................. 109
   4.4 Bibliographic Notes ....................................... 112

5 Congruences and Bisimulations 115
   5.1 Smooth Equivalence Relations ................................ 117
   5.2 Factoring .................................................. 131
   5.3 Bisimulations .............................................. 137
   5.4 2-Bisimulations ........................................... 145
   5.5 Simple Relations .......................................... 149
   5.6 Case Study: The Converse of a Stochastic Relation ............... 154
   5.7 Case Study: Simple Relations for Counting ...................... 165
   5.8 Bibliographic Notes ....................................... 170
### Contents

6 **Interpreting Modal and Temporal Logics**  
6.1 Modal Logics .................................................. 175  
6.2 Infinite Paths for Interpreting Temporal Logics ...................... 189  
6.3 Bisimulations for **CSL** ........................................ 201  
6.4 Bibliographic Notes ............................................. 212  

A **Measure Theory and Topology — A Refresher**  
A.1 Measurable Spaces ............................................... 215  
A.2 Polish and Analytic Spaces ...................................... 217  
A.3 Probability Measures ............................................ 221  

B **Notations etc.** ................................................. 229  
B.1 Categories ....................................................... 229  
B.2 Spaces .......................................................... 230  
B.3 Other .............................................................. 230  

**Bibliography** ....................................................... 232  

**Index** .............................................................. 239
Chapter 1

Introduction

A coalgebra for the endofunctor \( \mathcal{F} : C \rightarrow C \) is a pair \( \langle x, t \rangle \), where \( t : x \rightarrow \mathcal{F}(x) \) is a morphism in category \( C \). The study of coalgebras provides many interesting vistas into the landscape of (theoretical) computer science and algebra, as can be witnessed for example from the survey paper [77] by Rutten. Particularly interesting are the connections to modal logics [11], in which the power set functor \( \text{Pow} \) in the category of sets plays a prominent rôle. For example, a coalgebra \( \langle x, t \rangle \) for \( \text{Pow} \) may be identified with a relation on set \( x \).

Consider this example. A formula in a simple modal logic with \( A \) and \( AP \) as set of actions resp. atomic propositions is defined recursively through

\[
\phi ::= \top \mid p \mid \phi' \land \phi'' \mid \langle a \rangle \phi'
\]

Thus \( \top \) is a formula indicating \textit{truth}, each atomic proposition \( p \in AP \) is a formula, the conjunctions of two formulas is one, and whenever we have a formula, then prefixing it with \( \langle a \rangle \) for an action \( a \in A \) will yield a formula again. The intuitive meaning of \( \langle a \rangle \phi \) is “it is possible that \( \phi \) holds after action \( a \)” (just like the diamond is interpreted in ordinary modal logic as indicating possibility). An interpretation will take a set \( S \) of states, assign to each action \( a \in A \) a relation \( R_a \subseteq S \times S \), and to each atomic proposition \( p \in AP \) a subset \( L(p) \subseteq S \) of states. \( L(p) \) indicates the set of states in which \( p \) is valid.

The underlying Kripke model \( K = (S, AP, (R_a)_{a \in A}) \) holds these data as a container and is used for defining the semantics, which is done recursively through \((p \in AP, a \in A, \text{writing } R_a \text{ as } \rightarrow a):\)

\[
K, s \models \top \iff s \in S
\]

\[
K, s \models p \iff s \in L(p)
\]

\[
K, s \models \phi' \land \phi'' \iff K, s \models \phi' \text{ and } K, s \models \phi''
\]

\[
K, s \models \langle a \rangle \phi \iff K, s' \models \phi \text{ for some } s' \text{ with } s \rightarrow a s'
\]

Thus we have \( K, s \models \langle a \rangle \phi \) iff we can find a \( \rightarrow a\)-successor \( s' \) to \( s \) so that \( K, s' \models \phi \).

In terms of coalgebras, this Kripke structure can be seen essentially as a coalgebra \( \langle S, t \rangle \) for the functor that maps each set \( X \) to \( \text{Pow} \ (A \times X) \).

Now look at this: We have a system composed of two processors which may fail, but which may be repaired; the processors work independently and in parallel. The atomic propositions tell us which processor works, so that

\[
AP = \{\text{op0, op1, op2, op12}\},
\]
the actions either indicate the failure of a processor, saying which processor survives
(e.g., if both processors are operational, action \(\text{sw}1\) indicates that processor 2 fails and
processor 1 survives; if at least one processor operates, action \(\text{sw}0\) has the effect that all
processors fail), or indicate a repair action, telling us which processor has to be repaired, hence

\[ A = \{\text{sw}2, \text{sw}1, \text{sw}0, \text{rep}1, \text{rep}2, \text{rep}12\}. \]

Figure 1.1 gives a Kripke structure; each arrow has an action as a label, the label indicates,
too, to which relation the pair belongs.
In state \(A\), both processors are operational, in state \(D\), none is. State \(B\) has system 2
fail, etc. The map \(L\) is evident from the figure as well.

Put

\[
\phi_1 := \langle \text{rep}12 \rangle \langle \text{sw}2 \rangle \langle \text{sw}0 \rangle \text{op}0,
\]

\[
\phi_2 := \langle \text{rep}12 \rangle \langle \text{sw}1 \rangle \langle \text{sw}0 \rangle \text{op}0,
\]

then it is easy to see that both e.g. \(D \models \phi_1\) and \(D \models \phi_2\) hold, so state \(D\) cannot dis-\ntinguish between these formulas, so it is not really important whether processor 1 fails, or
processor 2. But probably it is.
Assume that probabilities are attached to the state transitions, as indicated in Figure 1.2
(to be precise, sub-probabilities, because the values do not add up to unity). Computing
the respective probabilities, it is clear that the probability for \(D\) accepting \(\phi_1\) is \(4.5 \cdot 10^{-2}\),
whereas for \(\phi_2\) it is \(18 \cdot 10^{-2}\), thus \(D\) can distinguish \(\phi_1\) and \(\phi_2\) now on the, say, 10% level.
Thus probabilities add to a more precise understanding of this system.
Addressing the question what kind of model might be adequate for probabilistic mod-\ning, one wants to pursue a development that runs in parallel to the coalgebraic one
above and considers coalgebras for the probability functor \(\mathcal{P}\). Thus a coalgebra for the
functor $S \mapsto \mathcal{P}(A \times S)$ would be replaced by a coalgebra for the functor $S \mapsto \mathcal{P}(A \times S)$ where $\mathcal{P}(A \times S)$ is the set of all probabilities over the set $A \times S$. Though straightforward, this approach raises some questions and needs to be refined. First, we have assumed implicitly that the sets we are dealing with are finite. If $A$ and $S$ are finite, so are $\mathcal{P}(A \times S)$, but $\mathcal{P}(A \times S)$ is an uncountable set as soon as $A \times S$ contains more than one element; so we leave the realm of finite sets with this construction. Thus it is conceptually difficult to iterate this construction, even for finite base sets; this, however, may be necessary when discussing monads. Moreover, some applications are intrinsically based on non-finite sets: consider e.g. a continuous time logic, where the residence times for the states may be non-negative real numbers. Second, when the universe may no longer assumed to be finite or countable, it may be difficult assigning probabilities to single elements in a standard environment (in contrast to a nonstandard one, see e.g. [60]), subsets are a more adequate domain.

But we know that we end up in considerable foundational difficulties when we assume that we assign a probability to each subset of an arbitrary set [31, 95]. Thus we need a structured subset of the power set as the domain for the probabilities, the structure being, as a student of measure theory knows, a $\sigma$-algebra. So we are poised to consider coalgebras for the functor that no longer takes an arbitrary set $X$ but rather a measurable space $(X, \mathcal{A})$ and assigns it all probabilities $\mathcal{P}(A \times X, \mathcal{H} \otimes \mathcal{A})$ on the measurable space $(A \times X, \mathcal{H} \otimes \mathcal{A})$ (since we assume that the actions are coming from a measurable space $(A, \mathcal{H})$ as well). Note that this switch entails changing the base category from the category of all sets with maps to the category of measurable spaces with measurable maps as morphisms.

But this is not yet enough. We want to model properties for these coalgebras that permit sensible applications like the study of bisimulations of various sorts or modelling probabilistically infinite paths in a logic for reactive systems. This is nearly hopeless
to do in general measurable spaces, because these spaces are not rich enough for supporting an interesting probabilistic structure. It is well known that Polish spaces, i.e., topological spaces that have a countable dense subset and for which a complete metric exists, provide enough support for the probability measures defined on their Borel sets (the smallest $\sigma$-algebra containing the open subsets) to permit the kind of constructions that we need. Examples of Polish spaces are countable discrete spaces (with the discrete topology), the real numbers $\mathbb{R}$, open or closed subsets of some Euclidean space, and even the measures on a Polish space under the weak topology; Polish spaces are closed under countable sums and products, and the general topological structure of Polish spaces has long been known very well, see [88, 51, 32, 73]. Thus if we have a Polish space $S$, e.g. the space $\text{PATHS}(S) := \prod_{i \in \mathbb{N}}(\mathbb{R}_+ \times S)$ of all infinite paths with timing information is Polish. Consequently we will work in the base category of all Polish spaces with Borel maps as morphisms, occasionally assuming continuity for the morphisms. Sometimes we will be able to extend the results to analytic spaces, so we can even go a step further and consider for most applications analytic spaces, hence measurable spaces that are the Borel images of Polish spaces.

Let us briefly return to the basic configuration of a coalgebra $\langle x, t \rangle$ with $t : x \to \mathbb{K}(x)$. Here the domain $x$ for the dynamic $t$ coincides with the domain for the functor $\mathbb{K}$. This is sometimes an uneasy restriction, so it is sometimes more adequate to work in a scenario that would look like $t : x \to \mathbb{K}(y)$. Separating the domain of the morphism $t$ from the functor's domain gives rise to a finer mode of description. For example a morphism $g : \langle x, t \rangle \to \langle y, s \rangle$ in the coalgebraic case is a morphism $g : x \to y$ with $s \circ g = \mathbb{K}(g) \circ t$, making this diagram commutative.

$$
\begin{array}{ccc}
x & \xrightarrow{g} & y \\
\downarrow t & & \downarrow s \\
\mathbb{K}(x) & \xrightarrow{\mathbb{K}(g)} & \mathbb{K}(y)
\end{array}
$$

The extended case requires a morphism $\langle x, y, t \rangle \to \langle a, b, s \rangle$ to be a pair $\langle g, h \rangle$ of morphisms $g : x \to a$ and $h : y \to b$ with $g \circ t = \mathbb{K}(h) \circ s$. This makes the diagram

$$
\begin{array}{ccc}
x & \xrightarrow{g} & a \\
\downarrow t & & \downarrow s \\
\mathbb{K}(y) & \xrightarrow{\mathbb{K}(h)} & \mathbb{K}(b)
\end{array}
$$

commutative. This observation permits separating the concerns of the domain from the codomain, which will be helpful in understanding some phenomena, as we will see.

Another instance where this separation of concerns pays is the description of congruences. In the coalgebraic case a congruence on $\langle x, t \rangle$ is essentially an equivalence relation on the carrier $x$ of the coalgebra that is compatible both with the dynamics $t$ and the
functor \( \mathfrak{F} \). In the extended case we deal with a pair of equivalence relations \((\alpha, \beta)\) that has to satisfy compatibility conditions with respect to both the dynamics and the functor. This separation will be of advantage in many places, for example when discussing Kripke models.

To summarize, we will discuss here morphisms of the kind \( x \rightarrow \mathfrak{F}(y) \) with \( \mathfrak{F} \) as the probability functor or one of its close relatives, defined usually on the category of Polish spaces. Again using the analogy to the power set functor we will see the objects we are dealing with as relations, albeit as stochastic ones. What we discuss will be outlined next.

**Contents**

We give a brief overview of the chapters’ contents, Figure 1.3 shows the dependencies between the sections.

**Chapter 1.** This introduction.

**Chapter 2.** This chapter will investigate categorial and probabilistic properties of stochastic relations. We take the properties of set-theoretic relations as a firm guide by looking at the power set functor on the category of sets with maps as morphisms. This functor assigns each set its power set, and it is the functorial part of the Manes monad (section 2.1). In parallel, we define the functor on measurable spaces that assigns each measurable space all its sub-probabilities. Under a suitable chosen \( \sigma \)-algebra, this is a measurable space again, so the functor is an endofunctor, and it can be seen as well as the functorial part of a monad, the Giry monad. It is shown that this is actually a monad; all this is done in section 2.2. These sections define and investigate the Kleisli construction for each monad, fairly basic for the discussions to follow. Since both Kleisli constructions have similar properties, we investigate in the case study in section 2.3 how these similarities can be exploited. A simple software architecture is modelled through a monad that enjoys a tensorial strength operator (both the Giry monad and the Manes monad belong to this family). In order to give a first impression what specifically can be done with stochastic relations, we also propose a probabilistic semantics for Ludwig, a simple imperative language with \texttt{while}-loop, in section 2.4. A partial correctness logic and semantics for Ludwig in terms of probabilistic relations are defined, correctness and completeness are established.

**Chapter 3.** The investigation of the sub-probability functor and the Giry monad would not be complete without having a look at the algebras that are associated with the monad. Mac Lane argues [62, Theorem VI.5.3] that the Kleisli construction and the Eilenberg-Moore algebras live at opposite ends of the spectrum of adjoint pairs of those functors which define a given monad, thus, having investigated the Kleisli construction, it is challenging to explore the other end. We identify the algebras for the Giry monad through positive convex structures in section 3.4, some illustrating examples are given in section 3.5. As always, the pair \((\mathfrak{S}(X), m_X)\) forms an algebra for each Polish space \( X \), where \( \mathfrak{S} \) is the sub-probability functor and \( m \) is the multiplication for the Giry monad. This example helps to identify in section 3.6 the left adjoint to the forgetful functor that
associates with each algebra the associated Polish space. As a case study for a possible application of the Eilenberg-Moore algebra we propose in section 3.7 derandomizations of Ludwig-programs and show that derandomization is correct with respect to the semantics in section 3.7; it is conjectured that derandomization is not complete.

This chapter lives in the category of Polish spaces with continuous maps. Transferring the results to the more general category of Polish spaces with Borel measurable maps (or even to the category of analytic spaces with Borel maps) is an open question.

Chapter 4. Many constructions in the theory of coalgebras depend on the assumption that the functor involved has at least weak pullbacks, see e.g. Rutten’s survey [77] for an overview. For practical purposes like the study of bisimulations it would be good to know whether or not the sub-probability functor has a weak pullback as well. These hopes are shattered rather quickly, so what one wants to look for is the existence of semi-pullbacks (thus each co-span of morphisms has a span, so that a commutative diagram is created). By formulating this problem as a selection problem and using the theory of measurable selectors, it is shown that in Polish spaces semi-pullbacks do exist. This solution can even be carried over to analytic spaces. Most work for a solution to this problem in section 4.3 concerns actually a measure extension problem — in probabilistic terms, a distribution on a sub-σ-algebra of the Borel sets has to be extended to the full Borel sets, see section 4.2. The problem is tackled using tools from classical analysis, based partially on the axiom of choice. It should be mentioned that this problem has been solved by A. Edalat [30] under the slightly restrictive assumption of universal measurability, using constructive techniques.

Chapter 5. Stochastic interpretations of modal logic show that equivalence relations are of interest that are countably generated, or, what amounts to the same, that are represented as the kernel of a Borel map. We study these relations that are called smooth in section 5.1. The interest in these relations stems in parts from the fact that factoring an analytic space with a smooth relation yields an analytic space again; this is well known [88, 51, 5]. We stress in particular the rôle of invariant Borel sets. They are important for two reasons: first, the inverse image of the Borel sets on the factor space under the factor map are just the invariant Borel sets. Second, this σ-algebra determines the equivalence relation uniquely (an observation that will be capitalized upon later, in section 6.3). These properties are investigated in greater detail in section 5.1, they are being made use of heavily for understanding congruences for stochastic relations. A pair $(\alpha, \beta)$ of smooth equivalence relations is called a congruence for a stochastic relation iff objects that cannot be separated through $\alpha$ and $\beta$ cannot be separated through the relation, see section 5.2. This section studies algebraic properties of congruences as well, an Isomorphism Theorem quite akin to the one in classical algebra can be established, providing the proper algebraic context.

Congruences are right at the heart of bisimulations. This is not really evident at first: bisimulations are introduced section 5.3 essentially through spans of morphisms in. This is a rather fruitful notion for investigating modal logics in chapter 6, but we have a look at immediate properties first, deriving a criterion for two stochastic relations to be bisimilar. Here we employ smooth equivalence relations and a technical condition which permits us to remotely investigate congruences on a stochastic relation, as if looking
from a vantage point. This leads to an intrinsic condition for bisimilarity (one looks only at the relations and does not have another, external, institution like a logic that assists in the decision) and has the interesting consequence that two relations are bisimilar provided they have isomorphic factor spaces. This criterion is also sufficient for compact metric spaces. The question remains open, however, whether or not this condition is sufficient in the general Polish or analytic case. A case study shows that bisimilarity does not break easily: we show that forming the converse of a stochastic relation — an interesting problem in itself — respects bisimilarity. If two relations are bisimilar, then their converses are as well; the properties of the converse relation are also explored, see section 5.6.

Bisimilarity may be specialized by taking projections as the morphisms involved; this is done in section 5.4 and leads to the notion of 2-bisimulation. Here the connections between bisimulations and congruences becomes fully visible: we show that each congruence can be used as the basis for a 2-bisimulation on a stochastic relation (when only one relation is involved). This is a fairly deep property that is derived again through technique from operations research using a suitable measurable selector. Once we know that, we are in a good position to tackle simple stochastic relations. These are those relations that have no nontrivial subsystems. It can be shown that they are completely characterized through injective Borel maps, providing a rather easy criterion for recognizing them (when encountering them in the street, say). As a consequence one derives that the sub-probability functor does not have a final system save for the case of proper probabilistic relations.

The interplay between bisimulations and simple systems is used in the theory of coalgebras for a calculus of coinduction, see [77, 78, 3]. Given the very simple structure of simple systems for the functor considered here, such an application does not seem to be realistic. Section 5.7 shows, however, that not all is lost. We indicate in this case study that the knowledge of simple systems permits at least some transfer results between discrete and continuous probability spaces when analyzing algorithms.

Chapter 6. This final chapter is devoted to stochastic interpretations of modal and continuous time logics. We extend the usual notion of modal logics for probabilistic purposes (incorporating into the language the notion of a probability with which a formula should be satisfied). This follows the trail of the seminal paper [59] by Larsen and Skou and takes up ideas from [20], in which Desharnais, Edalat and Panangaden go beyond the discrete approach discussed so far. It is then an easy undertaking to extend these ideas to general modal logics [11]. Because stochastic relations are available here in full generality, the interpretation needs not be confined to labelled Markov transition systems, rather a treatment of general Kripke models becomes feasible. This is proposed in section 6.1. The relationship between stochastic Kripke models and those based on set theoretic relations is investigated in this section as well, where we capitalize on the support function for rendering a nondeterministic Kripke model from a stochastic one. The relationship between bisimulations and the Hennessy-Milner equivalence relation [44] is scrutinized, it is shown how the characterization of bisimilar stochastic relations can be used as a basis for establishing that bisimilarity and Hennessy-Milner equivalence are really the same.

Whereas modal logics deals with finite paths, continuous stochastic logic is used to model
reactive systems, hence infinite paths have to be taken into account. These paths are usually alternating between states the system is in and residence times for indicating how long the system remains in this state before a state change occurs. Probabilistically, this is modelled through the projective limit of a process in which state changes and residence times are stochastically independent (this refers to one step in the system), see section 6.2. The logic, dubbed CSL, distinguishes state formulas (which display some sort of local behavior) from path formulas (which entertain long running properties). We discuss in section 6.3 another kind of bisimilarity: call two states $F$-bisimilar iff they satisfy exactly the same formulas from a given set $F$ of state formulas; since there is as usual a bridge between state formulas and path formulas, this notion of bisimilarity captures path formulas as well. Because $F$ is at most countable, this relation is smooth, hence it gives rise to a congruence on the interpreting relation. And here we are again: we can use the invariant sets of this smooth equivalence to determine sets $G$ for which $F$-bisimilarity and $G$-bisimilarity are identical. One wants these sets $G$ of course be as large as possible for maximizing the effect with minimal resources. It will help solving the problem of deciding whether for the set $AP$ of atomic propositions $AP$-bisimilarity is equal to $\mathcal{L}_{AP}$-bisimilarity, where $\mathcal{L}_{AP}$ is the set of all formulas. It is clear that this question is motivated through practically considerations: If the answer would be in the positive, one would have only to test the atomic propositions in order to make statements regarding the entire set of formulas. These questions are investigated in section 6.3; unfortunately, there is no clear cut, simple answer: as usual, it depends, in this case on the invariant sets. We finally define in this section the extension of a set $F$ of formulas as the set of all formulas that have the same invariant sets as $F$ and investigate an equivalence result involving 2-bisimulations.

Appendix. We work usually in a Polish space, sometimes in an analytic one, because most constructions which we propose cannot be undertaken in general measurable spaces, they demand a Polish structure. For example, a projective limit for an arbitrary projective system would not exist when not working in a Polish space, but a system like that is necessary when interpreting CSL. The Appendix offers a refresher in measure theory, Borel sets and Polish spaces. It is essentially taken from the standard sources quoted there; the book [88] by S. M. Srivastava has been particularly helpful.

Acknowledgements. Most of the work is based on the author’s research over the last couple of years, and it has been shaped by comments of many people, some of whom I do not even know, because they were the anonymous referees for conferences and journals. Prakash Panangaden provided some helpful and constructive insights, Michail Jerschov and Dieter Pumplün did substantially so. The Deutsche Forschungsgemeinschaft funded part of the research leading to this work through grants for the project Algebraische Eigenschaften stochastischer Relationen.

Writing of this book was done mostly during a sabbatical, part of which I was fortunate enough to spend in the colline pisane in the congenial atmosphere of Fritz-Hans Kaltenbach’s FATTORIA PUGNANO. I was permitted access to the impressive Computer Science library of the University of Pisa, which was made possible by Professor Carlo Montangero. Alla Stankjawitschene, my secretary, provided help and support in many things, Helmut Henning has given me patiently advice whenever I run into problems.
with typesetting. Georgios Lajios and Ignacio Viglizzo offered comments that did improve the representation. Stefan Dißmann's superb organizational skills made many things easier for me. I want to thank them all.

Bochum and Dortmund,                                                Ernst-Erich Doberkat
January 2006
Chapter 2

Categorial and Probabilistic Aspects of Stochastic Relations

Contents

2.1 The Manes Monad ........................................... 17
2.2 The Giry Monad ........................................... 19
  2.2.1 Adding a Monoid .................................... 21
  2.2.2 Stochastic Relations ................................ 24
2.3 Case Study: Architectural Modelling Through Monads. ....... 27
  2.3.1 A First Example .................................... 31
  2.3.2 First Steps .......................................... 32
  2.3.3 The Basic Construction .............................. 36
  2.3.4 Stratifying Graphs .................................. 39
  2.3.5 The General Case ................................. 41
  2.3.6 System Evolution .................................. 48
  2.3.7 Related Approaches ............................... 53
2.4 Case Study: Probabilistic Semantics of a Simple Language .... 55
  2.4.1 The Language: Ludwig .............................. 55
  2.4.2 Some Preparations and a Fixed Point .............. 56
  2.4.3 Partial Correctness Semantics ...................... 63
  2.4.4 A Partial Correctness Logic ....................... 64
  2.4.5 Consistency and Completeness ..................... 65
2.5 Bibliographic Notes ........................................ 67

We will investigate in this chapter the relationship between non-deterministic and stochastic relations from the point of view of category theory. The former relations are just sets of pairs, the latter may be interpreted as sets of pairs with a weight attached to it, the weight indicating the probability that the pair is selected. This is the intuitive and somewhat naive view which will be refined in this chapter. We show that the Kleisli construction provides a formal link between these kinds of relations. Selecting as a base category the category of sets and the power set functor, we will obtain non-deterministic relations through the Kleisli construction. Selecting the category of measurable spaces
and the sub-probability functor, we will obtain stochastic relations through this construction. They will be defined formally here as a result of these discussions. This is the rough picture, which will be refined somewhat. We will point out systematically some similarities. This is done through a discussion of the corresponding monads. When having a monad, one usually wants to know what the algebras for this monad looks like (because the algebras permit a reconstruction of an adjunction giving rise to that monad, just as the Kleisli construction does). We do this step as well for both monads, where the development for the monad based on the power set functor is well known, but for one based on the sub-probability it is not completely. It will be shown in chapter 3 what the algebras for the latter functor looks like, and here we obtain also an explicit characterization for an important special case. But the picture does not look that uniform: the algebras for the power set functor are just the $\sup$-complete orders, the algebras for the sub-probability functor are the positive convex structures.

We are dealing in this chapter mainly with categorical constructions, and we investigate the category of stochastic relations a bit more closely. We prepare for dealing with the problem of the existence of pullbacks, which then will be undertaken in chapter 4 in detail. This question in turn will be later of some significance, when we discuss bisimulations and their relations to modal and temporal logic. The problem is quite trivial for non-deterministic relations (you basically write the pullback down explicitly), but it is far too strongly posed for stochastic relations, even the request for weak pullbacks is not weak enough. We will show in chapter 4 that semi-pullbacks exist in the category of stochastic relations over analytic spaces, and that this is the most we can expect: no weak pullbacks usually exist, as an example shows. Reflecting this on the background of similarities between both kinds of relations, we see that constructions that are easily carried out for the set-theoretic case are undertaken with difficulties for the probabilistic case (if at all). We will encounter this phenomenon later on again. It suggests that a construction like an abstract specification of relations that can be interpreted sensibly both over non-deterministic and over stochastic relations may work in special cases, but may be difficult to pursue in general.

The present chapter introduces first the sub-probability functor on the category of measurable spaces. It investigates this functor, shows that it gives rise to a monad, has a look at the Kleisli product and identifies for a special case the algebras for this functor. Two case studies will provide some insight into possible applications.

To emphasize the similarities between non-deterministic and stochastic relations, we first look at the monad that is defined through the power set functor on $\text{Set}$. When modelling a software architecture in section 2.3, we will require an additional argument to the indication of the system’s work, and we assume for this purpose that this is modelled through a monoid $H$ with $1$ as an identity. Such a monoid could be a group, the free semigroup over an alphabet, or a $\vee$-semilattice with a smallest element. This additional argument will enter the constructions here at little additional cost but will give later an additional amount of flexibility.

This chapter contains two case studies as well, indicating the broad range of the concepts defined and investigated here. The first one deals with architectural modelling, the second one discusses the semantics of a simple programming language. It is shown that architectural models for a very popular software architecture may be formulated relationally. Since the basic mathematical demands for a relational model are the same for non-deterministic and for stochastic relations, the model is formulated sufficiently
2.1 The Manes Monad

general, so that both families of relations are covered (actually, the incorporation of the monoid permits tuning the functor a bit by incorporating additional information that may be used for bookkeeping and the like). Turning from programming in the large to programming in the small, hence from the discussion of the structure of a large software system to detailed issues of individual programs, and from general relations to the stochastic species, in the next case study the focus is laid on the semantics of a simple programming language. A probabilistic semantics is formulated, in particular the handling of iterative control structures using fixed points is studied (given that neither a convenient $Y$-operator, nor a complete partial order suggesting an application of the Kleene-Knaster-Tarski Theorem, nor a contractive map in a complete metric space proposing the use of Banach’s Theorem is available).

This chapter as a whole shows that the similarities between the families of relations considered are plentiful and interesting. They are translated from properties of the associated monads, in particular from the respective functors, which sit in their rôle as masterminds in the background and control the properties of their Kleisli products, sometimes remaining discreetly in the background, sometimes entering the bright sunlight through a direct argument.

**Monads.** Recall that in a category $\mathbf{C}$ an endofunctor $\mathcal{T}$ together with the natural transformations $\epsilon : I \mathbf{C} \to \mathcal{T}$ (the unit) and $m : \mathcal{T} \mathbf{C} \to \mathcal{T}$ (the multiplication) is a monad iff these diagrams commute

![Monad Diagram](https://via.placeholder.com/150)

(here $I\mathbf{C}$ is the identity on $\mathbf{C}$). The commutativity of the leftmost diagram is expressed for an object $x$ of $\mathbf{C}$ through

$$m_x \circ \mathcal{T}(m_x) = m_x \circ m_{\mathcal{T}(x)},$$

while the commutativity of the rightmost diagram is written down as

$$m_x \circ \epsilon_{\mathcal{T}(x)} = x = m_x \circ \epsilon(x).$$

These expressions are sometimes easier to handle than the purely functorial notation in the diagrams above.

**2.1 The Manes Monad**

The power set functor $\mathcal{P}\mathbf{ow}$ assigns to each set $X$ its power set $\mathcal{P}\mathbf{ow}(X)$, and assigns to each map $f : X \to Y$ the map $\mathcal{P}\mathbf{ow}(f) : \mathcal{P}\mathbf{ow}(X) \to \mathcal{P}\mathbf{ow}(Y)$, mapping $A \subseteq X$ to

$$\mathcal{P}\mathbf{ow}(f)(A) := f [A] := \{ f(x) \mid x \in A \}.$$
Define \( m : \mathcal{P}^2 \to \mathcal{P} \) through
\[
m_X : \mathcal{P} (\mathcal{P} (X)) \ni A \mapsto \bigcup A \in \mathcal{P} (X)
\]
and \( e : I \to \mathcal{P} (X) \) through
\[
e_X : X \ni x \mapsto \{ x \} \in \mathcal{P} (X).
\]
Elementary calculations show that both \( m \) and \( e \) form indeed natural transformations:

Let \( f : X \to Y \) be a map, then this diagram is commutative:

\[
\begin{array}{ccc}
\mathcal{P} (\mathcal{P} (X)) & \xrightarrow{m_X} & \mathcal{P} (X) \\
\mathcal{P} (\mathcal{P} (f)) & \xrightarrow{m_Y} & \mathcal{P} (f) \\
\mathcal{P} (\mathcal{P} (Y)) & \xrightarrow{m_Y} & \mathcal{P} (Y)
\end{array}
\]

In fact, if \( A \in \mathcal{P} (\mathcal{P} (X)) \), then
\[
(m_Y \circ \mathcal{P} (\mathcal{P} (f))) (A) = \bigcup \{ f [x] \mid x \in A \} = (\mathcal{P} (f) \circ m_X) (A).
\]

It is also not difficult to establish that \( (\mathcal{P}, e, m) \) satisfies the laws of a monad; this monad will be referred to as the \textit{Manes monad}.

Augmenting the construction by adding semigroup \( H \), we define
\[
\mathcal{M} (X) := \mathcal{P} (H \times X),
\]
and if \( f : X \to Y \) is a map, \( A \subseteq H \times X \), then
\[
\mathcal{M} (f) (A) := \{ \langle h, f(x) \rangle \mid \langle h, x \rangle \in A \}
\]
defines the action of the functor on the morphisms of Set. Now define
\[
m_X : \mathcal{M} (\mathcal{M} (X)) \to \mathcal{M} (X)
\]
upon setting
\[
m_X (A) := \bigcup \{ \langle h_1 h_2, x \rangle \mid \langle h_2, x \rangle \in b \},
\]
then an easy but somewhat space consuming calculation reveals that \( m : \mathcal{M}^2 \to \mathcal{M} \) is a natural transformation. The natural transformation \( e : \mathcal{I}_{\text{Set}} \to \mathcal{M} \) is defined by \( e_X : x \mapsto \{ \langle 1, x \rangle \} \). A standard calculation shows then that \( (\mathcal{M}, e, m) \) is a monad in Set.

Let \( (\mathcal{I}, e, m) \) be a monad in category \( C \), then a \textit{Kleisli morphism} \( f : a \rightsquigarrow b \) between objects \( a \) and \( b \) is a morphism \( f : a \to \mathcal{I} b \). The composition \( g * f \) of Kleisli morphisms \( f : a \rightsquigarrow b \) and \( g : b \rightsquigarrow c \) is defined through
\[
g * f := m_c \circ \mathcal{I} g \circ f,
\]
2.2 The Giry Monad

where \( \circ \) is the composition in \( C \). The properties of a monad take care of associativity and the fact that the identity morphism in the original monad gives rise to an identity for the Kleisli composition. It is well known that \( C \) forms a category with this composition [62, Theorem VI.5.1], the Kleisli category for the monad.

A Kleisli morphism between sets \( X \) and \( Y \) is a relation between \( X \) and \( H \times Y \). This is well investigated for the case that the monoid \( H \) is trivial, cf. [9, 16.1.4]; it generalizes to the present case. Let \( R : X \to \mathcal{M}(Y) \) and \( S : Y \to \mathcal{M}(Z) \) be Kleisli morphisms, thus we may either see \( R \) and \( S \) as maps to the corresponding power sets, or we interpret \( R \subseteq X \times (H \times Y) \) and \( S \subseteq Y \times (H \times Z) \) as relations; both views will be made use of interchangeably, depending on the convenience of use.

The (Kleisli-) product of \( R \) and \( S \) is identified in Proposition 2.1.1 which summarizes this example.

**Proposition 2.1.1** \( (\mathcal{M}, \epsilon, m) \) is a monad in the category \( Set \). The Kleisli product for the relations \( R \subseteq X \times (H \times Y) \) and \( S \subseteq Y \times (H \times Z) \) is given through

\[
(S \ast R)(x) = \{ \langle h, z \rangle | \exists y \in Y : \langle h, y \rangle \in R(x) \land \langle h, z \rangle \in S(y) \}.
\]

Assume that the semigroup \( H \) is trivial, then functor \( \mathcal{M} \) equals the power set functor \( \mathcal{P} \omega \), the natural transformations \( m \) and \( \epsilon \) are adjusted as well. We obtain as a consequence of Proposition 2.1.1 the well-known version we will usually work with:

**Corollary 2.1.2** \( (\mathcal{P} \omega, \epsilon, m) \) is a monad in the category \( Set \). The Kleisli product for the relations \( R \subseteq X \times Y \) and \( S \subseteq Y \times Z \) is given through

\[
(S \ast R)(x) = \{ z | \exists y \in Y : y \in R(x) \land z \in S(y) \}.
\]

\( \dashv \)

\( 2.2 \) The Giry Monad

We will investigate now the probabilistic counterpart to the Manes monad. It assigns to each measurable space its sub-probability measures. We demonstrate first that this constitutes a functor, then we will show that it is actually the functorial part of a monad, the Giry monad, investigated first by M. Giry [39].

Assume that \( X \) and \( Y \) are measurable spaces.\(^1\) Let \( f : X \to Y \) be a measurable map, then \( f \) induces a map \( \mathcal{S}(f) : \mathcal{S}(X) \to \mathcal{S}(Y) \) upon setting (\( \mu \in \mathcal{S}(X), B \subseteq Y \) measurable)

\[
\mathcal{S}(f)(\mu)(B) := \mu(\{ x \in X | f(x) \in B \}) = \mu(f^{-1}[B]).
\]

\( \mathcal{S}(f)(\mu) \) is referred to as the image measure of \( \mu \) under \( f \). Integration with respect to the image measure may be captured through the Change of Variable formula which will be somewhat helpful in the sequel.

\(^1\)Unless there is the danger of ambiguity, we will omit the notation of the \( \sigma \)-algebras from now on. They are assumed to be fixed. The following conventions will be adhered to: The space of sub-probabilities will be endowed with the corresponding weak-*\( \sigma \)-algebra (see section A.3.1 for a definition), products will carry the product \( \sigma \)-algebra of their factors, and Polish or analytic spaces will always have their Borel sets, unless otherwise indicated.

Derivations from these conventions will be made explicit.
**Proposition 2.2.1 (Change of Variables)** Let \( g \in \mathcal{F}(Y) \) be a bounded and measurable function, then
\[
(\dagger) \quad \int_Y g(y) \mathcal{G}(f)(\mu)(dy) = \int_X (g \circ f)(x) \mu(dx).
\]

**Proof** (Sketch) We have a look at all \( g \) for which the assertion is true:
\[
\mathcal{F}_0 := \{ g \in \mathcal{F}(Y) \mid (\dagger) \text{ holds for } g \}
\]

Then \( \chi_B \in \mathcal{F}_0 \), provided \( B \subseteq Y \) is measurable. This is so since
\[
\int_Y \chi_B \, d\mathcal{G}(f)(\mu) = \mathcal{G}(f)(\mu)(B) = \mu(f^{-1}[B]) = \int_X \chi_B \circ f \, d\mu.
\]

It is clear from the integral’s additivity that \( \mathcal{F}_0 \) is a linear space, so that measurable step functions are contained in \( \mathcal{F}_0 \). Since for each \( g \in \mathcal{F}(Y) \) with \( g \geq 0 \) there exists an increasing sequence \( (g_n)_{n \in \mathbb{N}} \) of measurable step functions such that \( g = \sup_{n \in \mathbb{N}} g_n \), we obtain
\[
\int_Y g \, d\mathcal{G}(f)(\mu) = \lim_{n \to \infty} \int_Y g_n \, d\mathcal{G}(f)(\mu) = \lim_{n \to \infty} \int_X g_n \circ f \, d\mu = \int_Y g \circ f \, d\mu.
\]

Consequently, \( \mathcal{F}_0 \) contains each nonnegative measurable and bounded function. Since each function \( g \) can be written as \( g = \max(g, 0) + \min(g, 0) \), it follows that \( \mathcal{F}_0 = \mathcal{F}(Y) \), hence the assertion is true for all measurable and bounded functions on \( Y \). 

The reader is probably more familiar with a version that permits changing real variables. It says that for a monotone and continuous differentiable maps \( g \) with domain \([a,b]\) and range \([\alpha,\beta]\) the equality
\[
\int_\alpha^\beta f(y) \, dy = \int_a^b (f \circ g)(x) \cdot |g'(x)| \, dx
\]
holds, whenever \( f \) is integrable over \([\alpha,\beta]\). This is the classical version of Calculus, and it is in fact a special case of the Proposition above. This is discussed at length in [45, Chapter 20.2].

Given a measurable map \( f \), the induced map \( \mathcal{G}(f) : \mathcal{G}(X) \to \mathcal{G}(Y) \) is measurable with respect to the corresponding weak-*-\( \sigma \)-algebras (section A.3.1). This is seen as follows: we have to establish \( \mathcal{G}(f)^{-1}[B^\star] \subseteq A^\star \), hence have to show that for each \( B^\prime \in B^\star \) its inverse image under \( \mathcal{G}(f) \) is in \( A^\star \). Here \( A \) resp. \( B \) are the \( \sigma \)-algebras on \( X \) and \( Y \). The construction of the weak-*-\( \sigma \)-algebra entails that we may assume \( B^\prime = \{ \nu \in \mathcal{G}(Y) \mid \nu(B) < t \} \) for some measurable \( B \subseteq Y, t \in \mathbb{R} \), since sets of this form generate the weak-*-\( \sigma \)-algebra \( B^\star \). But then
\[
\mu \in \mathcal{G}(f)^{-1}[B^\prime] \iff \mu \in \{ \mu' \in \mathcal{G}(X) \mid \mu'(f^{-1}[B]) < t \},
\]
because of the assumption on \( f \)'s measurability, \( f^{-1}[B] \subseteq X \) is measurable. This establishes the measurability of \( \mathfrak{G}(f) \). We have shown:

**Proposition 2.2.2** \( \mathfrak{G} \) is an endofunctor on the category \( \text{Meas} \) of measurable spaces with measurable maps as morphisms. \( \dashv \)

### 2.2.1 Adding a Monoid

We will now define the Giry monad on \( \text{Meas} \), and similar to the discussion of the Manes monad we will do this a bit more general by taking a monoid \( H \) into account. We assume \( H \) being endowed with a \( \sigma \)-algebra \( \mathcal{H} \) which makes multiplication measurable, when \( H \times H \) carries the product \( \sigma \)-algebra \( \mathcal{H} \otimes \mathcal{H} \).

Examples for such measurable monoids are given by topological monoids; the Borel sets then form the canonical measurable structure. Topological groups are probably the most prominent examples. If \( H \) is a \( \lor \)-semilattice with a smallest element, then it is not difficult to see that \( \lor \) is a continuous operation, when \( H \) is endowed with the interval topology (i.e. the topology which has open intervals as subbase). Taking again the Borel sets for this topology, we see that these semilattices yield measurable monoids, too.

We state as a preparation for the definition of the monad's multiplication:

**Lemma 2.2.3** Let \( f : X \to Y \) be a measurable map, and assume that \( C \) is a measurable subset of \( H \times Y \). Then \( \Gamma_C(\nu, s) := \nu(\{ \langle t, x \rangle \mid \langle st, f(x) \rangle \in C \}) \) is a real-valued measurable map on \( \mathfrak{G}(H \times X) \times H \).

**Proof**

1. Consider the set \( C \) of all members \( C \) of \( H \otimes \mathcal{B} \) for which \( \Gamma_C \) has the desired property. We will analyze the properties of \( C \) now with the goal of showing that \( C \) equals \( H \otimes \mathcal{B} \).
2. Assume that \( C = G \times D \) with \( G \in \mathcal{H} \) and \( D \in \mathcal{B} \). Since the semigroup multiplication is measurable, we know that \( G' := \{ \langle g, h \rangle \mid gh \in G \} \) is an element of \( \mathcal{H} \otimes \mathcal{H} \), so that

\[
\{ \langle t, x \rangle \mid \langle st, f(x) \rangle \in G \times D \} = G'_s \times f^{-1}[D].
\]

(recall that \( G'_s \) is the vertical cut of \( G' \) at \( s \), see section A.3.2, and that \( f^{-1}[D] \subseteq X \) is measurable). Hence

\[
\Gamma_{G \times D}(\nu, s) = \nu(G'_s \times f^{-1}[D]).
\]

An argument very similar to that establishing Lemma A.3.5 shows that \( (\nu, s) \mapsto \nu(Q_s \times W) \) is \( (\mathcal{H} \otimes \mathcal{A})^\ast \otimes \mathcal{H} \)-measurable, whenever \( Q \in \mathcal{H} \otimes \mathcal{H} \) and \( W \in \mathcal{A} \). This implies that \( \Gamma_{G \times D} \) constitutes a measurable map, so that all measurable rectangles are members of \( C \). The measurable rectangles form a family of sets which is closed under finite intersections.
3. Because sub-probabilities are finitely and countably additive, we have

\[
\Gamma_{(H \times X) \setminus C}(\nu, s) = \nu(H \times X) - \Gamma_C(\nu, s),
\]

thus \( C \) is closed under complementation, and

\[
\Gamma_{\bigcup_{n \in \mathbb{N}} C_n}(\nu, s) = \sum_{n \in \mathbb{N}} \Gamma_{C_n}(\nu, s),
\]
whenever \((C_n)_{n \in \mathbb{N}}\) is a disjoint sequence, thus \(C\) is closed under disjoint countable unions. From the \(\pi\)-\(\lambda\)-Theorem (Theorem A.1.1) we infer now that \(C\) contains all sets \(C \in \mathcal{H} \otimes B\). This establishes the claim. \(\dashv\)

Lemma 2.2.3 states that \(\Gamma_C\) is jointly measurable in both arguments. This entails that we may use \(\Gamma_C\) as an integrand, given a sub-probability on its domain for integration. It has also as a consequence that upon fixing one argument, the arising partial map is measurable, so that \(\Gamma_C\) is measurable separately in each variable (it is well known that joint measurability is strictly stronger than separate measurability in each variable). We will make use of this observation as well.

Now define after these somewhat lengthy preparations for \(x \in X\), the measurable subset \(A \subseteq H \times X\) and the measure \(\mu \in \mathcal{G}(H \times \mathcal{G}(H \times X))\) the functor \(\mathcal{G}\), unit \(\epsilon\) and multiplication \(m\) through

\[
\mathcal{G}(X) := \mathcal{G}(H \times X, \mathcal{H} \otimes A) \\
\epsilon_X(x) := \delta_{(1,x)} \\
m_X(\mu)(A) := \int_{H \times \mathcal{S}(H \times X)} \nu(\{(h, x) \mid (gh, x) \in A\}) \mu(d(g, p))
\]

Lemma 2.2.3 tells us that the integrand for the definition of \(m_X\) is actually measurable. If \(f : X \rightarrow Y\) is measurable, we put

\[
\mathcal{G}(f)(\mu)(B) := \mu(\{(h, x) \mid (h, f(x)) \in B\}) = \mu((id_H \times f)^{-1}[B]) = \mathcal{G}(id_H \times f)(\mu)(B).
\]

Consequently, \(\mathcal{G}(f) : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)\) is measurable, and if \(\psi \in \mathcal{F}(H \times Y)\), the Change of Variable formula (Proposition 2.2.1) implies that

\[
\int_{H \times Y} \psi(h, y) \mathcal{G}(f)(\mu)(d(h, y)) = \int_{H \times X} \psi(h, f(x)) \mu(d(h, x))
\]

holds.

We are now in a position to show that \((\mathcal{G}, \epsilon, m)\) is a monad in \(\text{Meas}\), adapting and extending Giry’s proofs [39] to the situation at hand.

**Lemma 2.2.4** \(\mathcal{G}\) is an endofunctor in \(\text{Meas}\), \(\epsilon : \mathcal{I}_{\text{Meas}} \rightarrow \mathcal{G}\) and \(m : \mathcal{G}^2 \rightarrow \mathcal{G}\) are natural transformations.

**Proof**

1. It is immediate that \(\mathcal{G} : \text{Meas} \rightarrow \text{Meas}\) is a functor, and that \(\epsilon\) is a natural transformation.

2. Let \(f : X \rightarrow Y\) be a measurable map, then we know that for \(\mu \in \mathcal{G}(Y)\) and for the measurable subset \(B \subseteq H \times Y\) these equations hold

\[
(m_Y \circ \mathcal{G}^2 f)(\mu)(B) = \int_{H \times \mathcal{S}(H \times Y)} (\mathcal{G} f)(q)(\{(h, y) \mid (gh, y) \in B\}) \mu(d(s, q)) = \int_{H \times \mathcal{S}(H \times X)} q(\{(h, x) \mid (gh, f(x)) \in B\}) \mu(d(s, q)).
\]
Again, an appeal to Lemma 2.2.3 makes sure that we are permitted to compute the integral, since the integrand constitutes a bounded function that is measurable jointly in both variables. The latter expression coincides with $(\mathcal{G} f \circ m_X)(\mu(B))$. Thus we have established that 
\[ (m_Y \circ \mathcal{G}^2 f) = (\mathcal{G} f \circ m_X) \]
holds. Consequently, $\mu : \mathcal{G}^2 \to \mathcal{G}$ is a natural transformation. \(\dashv\)

This is a preparation for establishing:

**Proposition 2.2.5** $\langle \mathcal{G}, e, m \rangle$ is a monad in $\text{Meas}$.

**Proof**

1. We need to demonstrate that both the associative and the unit laws hold. The Change of Variable formula implies that 
\[ \int_{G \times \mathcal{G}(X)} \psi d(\mathcal{G} e_X)(\mu) = \int_{H \times X} \psi(h, e_X)(x) p(d(h, x)), \]
whenever $\mu \in \mathcal{G}(X)$, and $\psi \in \mathcal{F}(H \times \mathcal{G}(X))$ is measurable and bounded. Consequently, 
\[ (m_X \circ \mathcal{G} e_X)(\mu)(B) = \int_{H \times X} e_X(\{g, x\} : \langle hg, x \rangle \in B) \mu(d(h, x)) = \mu(B) = (m_X \circ e_{\mathcal{G}(X)})(\mu)(B) \]
is true for every $\mu \in \mathcal{G}(X)$, and for every measurable subset $B$ of $H \times X$. This establishes the unit laws.

2. As far as the associative law is concerned, fix $r \in \mathcal{G}^3(X)$, and a measurable subset $E$ of $H \times \mathcal{G}(X)$. The Change of Variable formula implies that 
\[ (m_X \circ m_{\mathcal{G}(X)})(r)(E) = \int_{H \times \mathcal{G}(X)} q(\{g, y\} : \langle hg, y \rangle \in E) m_{\mathcal{G}(X)}(r)(d(h, q)) = \int_{H \times \mathcal{G}^2(X)} \left( \int_{H \times \mathcal{G}(X)} q(\{j, y\} : \langle gj, y \rangle \in E) p(d(h, q)) \right) r(d(g, p)) \]
On the other hand, expanding the definitions, and applying the Change of Variables formula suitably, it is seen that these transformations hold:
\[ (m_X \circ \mathcal{G} m_X)(r)(E) = \int_{H \times \mathcal{G}(X)} p(\{h, y\} : \langle gh, y \rangle \in E) \mathcal{G} m_X(r)(d(g, p)) = \int_{H \times \mathcal{G}^2 X} m_X(q)(\{h, y\} : \langle gh, y \rangle \in E) r(d(g, q)) = (m_X \circ m_{\mathcal{G}(X)})(r)(E). \]
This shows that the associative law is valid. \(\dashv\)

We identify the product in the Kleisli category associated with this monad:
Proposition 2.2.6 Let $X, Y$ and $Z$ be measurable spaces. Assume that $K : X \rightsquigarrow Y$ and $L : Y \rightsquigarrow Z$ are Kleisli morphisms for the monad $\langle \mathcal{S}, \epsilon, m \rangle$. Then the Kleisli product $L \ast K$ for $K$ and $L$ is given through

$$
(L \ast K)(x)(C) = \int_{H \times Y} L(y)(\{\langle h, x \rangle | \langle gh, x \rangle \in C\}) K(x)(d\langle g, y \rangle).
$$

Proof 1. Let $C$ be a measurable subset of $H \times Z$, then the definition of the Kleisli product yields

$$
(L \ast K)(x)(C) = (m_Z \circ \mathcal{S} (L) \circ K)(x)(C)
$$

$$
= m_Z ((\mathcal{S} (L) \circ K)(x))(C)
$$

$$
= \int_{H \times \mathcal{S}(Z)} \mu(\{\langle t, z \rangle | \langle st, z \rangle \in C\}) (\mathcal{S} (L) \circ K)(x)(d\langle s, \mu \rangle).
$$

2. If $\psi \in F(H \times \mathcal{S}(Z))$ and $\mu \in \mathcal{S}(Y)$, the Change of Variables formula implies that

$$
\int_{H \times \mathcal{S}(Z)} \psi d\mathcal{S} (L)(\mu) = \int_{H \times Y} \psi(t, L(y)) \mu(d\langle t, y \rangle).
$$

Inserting this into the equation above, the result follows. ⊣

2.2.2 Stochastic Relations

Summarizing the discussion for a trivial monoid, we obtain as an extension to Proposition 2.2.2 the following Corollary. It is stated separately because we will use it over and over again.

Corollary 2.2.7 $\langle \mathcal{S}, \epsilon, m \rangle$ is a monad in the category $\text{Meas}$ of measurable spaces with measurable maps as morphisms. Assume that $K : X \rightsquigarrow Y$ and $L : Y \rightsquigarrow Z$ are Kleisli morphisms for this monad. Then the Kleisli product $L \ast K$ for $K$ and $L$ is given through

$$
(L \ast K)(x)(C) = \int_Y L(y)(C) K(x)(dy).
$$

⊣

We will investigate these Kleisli morphisms in greater detail: they constitute just the stochastic relations. The name suggests the similarity with set-theoretic (or non-deterministic) relations.

Definition 2.2.8 A stochastic relation $K = (X, Y, K)$ between the measurable spaces $X$ and $Y$ is a Kleisli morphism $K : X \rightsquigarrow Y$ for the monad $\langle \mathcal{S}, \epsilon, m \rangle$.

In the literature on Probability Theory (e. g., [86, 61, 10, 73, 88, 51]), a stochastic relation is called a transition sub-probability or a sub-Markov kernel. We want, however, to emphasize the mutual relationship among these relations, so we stick to this name which is closer to the Computer Science point of view and to the to applications there. This is another, easier to handle characterization of stochastic relations.

24
Proposition 2.2.9  The following statements are equivalent for measurable spaces $X$ and $Y$:

1. $K : X \rightsquigarrow Y$ is a stochastic relation.

2. $K : X \times B \to [0,1]$ is a map such that

   (a) $x \mapsto K(x)(B)$ is measurable for each $B \in B$,

   (b) $K(x) \in \mathcal{G}(Y)$ for each $x \in X$.

Here $B$ is the $\sigma$-algebra on $Y$.

Proof 1. Suppose that $K$ is a stochastic relation, then $K : X \to \mathcal{G}(Y)$ is a measurable map. The definition of the weak-*-$\sigma$-algebra implies that the evaluation map $x \mapsto K(x)(B)$ is measurable for each $B \in B$.

2. Assume conversely that $K$ has the properties from part 2. It is clear that $K$ maps $X$ to $\mathcal{G}(Y)$, so measurability has to be established. Again, this follows readily from the definition of the weak-*-$\sigma$-algebra. \(\square\)

Corollary 2.2.10  Assume that $K : X \rightsquigarrow Y$ is a stochastic relation, and $D \in A \otimes B$, where $A$ and $B$ are the respective $\sigma$-algebras. Then $x \mapsto K(x)(D_x)$ is a measurable map.

Proof  Consider $D := \{D \in A \otimes B \mid x \mapsto K(x)(D_x) \text{ is measurable}\}$. Evidently both $\emptyset$ and $X \times Y$ are members of $D$, and since we can calculate $D_x$ for $D = A \times B$ as

$$D_x = \begin{cases} B, x \in A, \\ \emptyset, x \notin A, \end{cases}$$

we see that all measurable rectangles are members of $D$; this generator is closed under finite intersections. If $D \in D$, then $((X \times Y) \setminus D)_x = Y \setminus D_x$, and $K(x)(Y \setminus D_x) = K(x)(Y) - K(x)(D_x)$. Consequently, $D$ is closed under complementation. Similarly, if $(D_n)_{n \in \mathbb{N}} \subseteq D$ is a sequence of disjoint sets, then, since $\bigcup_{n \in \mathbb{N}} D_n)_x = \bigcup_{n \in \mathbb{N}} (D_n)_x$, and since the infinite sum of a sequence of measurable functions is measurable again, provided it exists (which it does in this case), we may conclude that $\bigcup_{n \in \mathbb{N}} D_n \in D$. Thus $D$ is closed under complementation and disjoint unions, and it contains a generator that is closed under finite intersections. From the $\pi$-$\lambda$-Theorem A.1.1 we see that $D = \sigma(\{A \times B \mid A \in A, B \in B\})$, hence the assertion is true for all product measurable sets. \(\square\)

Discussion.  Proposition 2.2.9 supports the view that a stochastic relation models randomly changing phenomena. Assume first that $K : S \rightsquigarrow S$ is a stochastic relation on a state space $S$ for some system. If the system is in state $s \in S$, then $K(s)(T)$ is interpreted as the probability that the system will change its state to a member of the measurable set $T \subseteq S$. Second, assume that $X$ and $Y$ are interpreted as the spaces of inputs and outputs of some randomly operating device. Then the value $K(x)(B)$ for a stochastic relation $K : X \rightsquigarrow Y$ is interpreted as the probability for an output to be a member of the measurable set $B \subseteq Y$ after the system has received input $x \in X$. Models like this are particularly attractive when outputs come from an uncountable set: here it is not always...
reasonable to assign to each individual \( y \in Y \) a probability, because such a probability needs usually to be zero. On the other hand it appears sensible to assign sets of outputs the probability to be involved. It should be mentioned that methods of nonstandard analysis [60, 52] try to balance these seemingly irreconcilable points of view.

Another point worth mentioning is that not for all inputs \( x \) the probability \( K(x)(Y) \) that an output will be delivered needs to be unity (for otherwise we would have postulated that \( K \) maps \( X \) to \( \mathcal{P}(Y) \) rather than to \( \mathcal{S}(Y) \)). This permits modelling systems that may encounter situations in which no output at all will be given, e.g., because the computation leading to an output does not terminate, see [69].

Suppose that the base spaces \( X \) and \( Y \) are identical, then a stochastic relation may be interpreted as a coalgebra for the sub-probability functor \( \mathcal{S} \), see e.g. [77]. This is of interest when modelling state transitions as hinted at above. The coalgebraic point of view appears quite attractive structurally, because it suggests to fit stochastic relations tightly under the roof of coalgebras, making tried and tested approaches available for investigating problems of stochastic relations. Unfortunately, this route can only be followed with partial success. There are two reasons for this: First, we will see that the sub-probability functor has some idiosyncratic properties making work with it sometimes a little strenuous (for example, due to the lack of weak pullbacks, see Proposition 4.3.6). Second, a coalgebra \( (x, c) \) for functor \( \mathcal{G} \) is defined as a morphism \( c : x \rightarrow \mathcal{G}x \), so the codomain of morphism \( c \) is just the image of its domain under \( \mathcal{G} \). This is rather restrictive, both structurally and regarding applications. “Unfolding” domain and codomain into two independent objects provides much needed maneuverability, as we will experience almost everywhere.

**Morphisms.** Given two stochastic relations \( K_1 \) and \( K_2 \) with \( K_i : (X_i, A_i) \rightharpoonup (Y_i, B_i) \ (i = 1, 2) \) a morphism \( f : K_1 \rightarrow K_2 \) is composed of two maps \( \phi : X_1 \rightarrow X_2 \) and \( \psi : Y_1 \rightarrow Y_2 \). Both maps should be measurable, and we will assume that both maps are onto. This is due to the observation that in the target system each element should be traced back to an element in the source system (so that there is no overabundance of elements in \( K_2 \) relative to \( K_1 \)). We formulate as the compatibility condition relating the probabilistic structures \( K_1 \) and \( K_2 \) that

\[
K_1(x_1)(\psi^{-1}[B_2]) = K_2(\phi(x_1))(B_2)
\]

holds for each \( x_1 \in X_1 \) and each \( B_2 \in B_2 \). Staying with the input-output model, we postulate that the probability of answering with an element of \( B_2 \) after input \( \phi(x_1) \) equals the probability of answering after input \( x_1 \) with an element which will be mapped by \( \psi \) to \( B_2 \) (thus with an element of \( \psi^{-1}[B_2] \)). The equation above may be reformulated as

\[
\mathcal{G}(\psi) \circ K_1 = K_2 \circ \phi,
\]

(composition \( \circ \) denoting composition of maps), and this leads to the following fundamental definition:

**Definition 2.2.11** Given two stochastic relations \( K_i : (X_i, A_i) \rightharpoonup (Y_i, B_i) \ (i = 1, 2) \), a morphism \( f : K_1 \rightarrow K_2 \) is a pair \( f = (\phi, \psi) \) of surjective maps such that

1. \( \phi : X_1 \rightarrow X_2 \) is \( A_1 \)-\( A_2 \)-measurable,
2. \( \psi : Y_1 \rightarrow Y_2 \) is \( B_1 \)-\( B_2 \)-measurable,
2.3 Case Study: Architectural Modelling Through Monads.

3. the diagram

\[
\begin{array}{ccc}
X_1 & \xrightarrow{\phi} & X_2 \\
\downarrow K_1 & & \downarrow K_2 \\
\mathcal{S}(Y_1, B_1) & \xrightarrow{\mathcal{S}(\psi)} & \mathcal{S}(Y_2, B_2)
\end{array}
\]

is commutative.

This is just a morphism in the comma category \((\mathcal{C}_{Meas'} \downarrow \mathcal{S})\) with \(\mathcal{C}_{Meas'}\) as the identity functor on the category Meas', which in turn is the subcategory of Meas that has surjective measurable maps as morphisms. Comma categories are introduced and discussed in [62, Chapter II.6].

**Listing Some Categories.** We will usually not work in the very general category of all measurable spaces but restrict ourselves to some more specialized base categories like BPol, the category of Polish spaces with Borel measurable maps as morphisms or the category Anl of analytic spaces, also with Borel measurable maps as morphisms. The category Stoch has stochastic relations \(K = (X, Y, K)\) for measurable spaces \(X, Y\) as objects and pairs of surjective maps according to Definition 2.2.11 as morphisms. Stochastic relations will be investigated also over these more specialized categories, and morphisms between stochastic relations are available there as well. PolStoch and anStoch denote the category of all stochastic relations over BPol resp. Anl. The objects of PolStoch will sometimes be called Polish objects, and, accordingly, analytic objects will be the objects in anStoch.

2.3 Case Study: Architectural Modelling Through Monads.

Nondeterministic and stochastic relations are for some problems really instances of the same relational phenomenon: we need little beyond the corresponding monads, and the relevant aspects of the applications will be taken care of through the Kleisli morphisms for that monad. This basic observation lies at the heart of Moggi's \(\lambda\)-calculus as well [67, 66], but in contrast to Moggi's work we are interested here in exploring the commonalities and the differences of two very specific monads. We will investigate the problem of modelling a popular, simple software architecture with this approach: given a monad with some additional features, we investigate modelling this architecture, and we show that both the Manes monad and the Giry monad are instances of it. It will be argued at the end that, since the architecture is so simple, this uniform approach of modelling is successful, and that a slightly more complicated software architecture will need more sophisticated categorial properties [57].

A pipeline is a popular architecture which connects computational components (filters) through connectors (pipes) so that computations are performed in a stream like fashion. The data are transported through the pipes between filters, gradually transforming inputs to outputs. This kind of stream processing has been made popular through UNIX pipes that serially connect independent components for performing a sequence of tasks.
Because of its simplicity and its easy to grasp functionality it is a pet architecture for demonstrating ideas about formalizing the architectural design space (not unlike the data type Stack for algebraic specifications or abstract data types). We will show in this section how to formalize this architecture in terms of monads, hereby including specifications through set theoretic or probabilistic relations as special cases.

**Software Architectures.** The structural aspects of a large programming system are captured through its (software) architecture. Initially, this term was used rather loosely, work being done during the 1990s in particular by M. Shaw and her associates have established a body of knowledge in the software engineering community about methods for structuring large systems. This translates into practical tools like architectural design languages.

An architecture for a system separates computation from control on the system’s level; while the former is represented by algorithms formulated in a programming language, the latter is formulated in terms of components (which carry out the computations) and connectors (which transport data from one component to another one). Connectors are elevated to first class rank making it possible to reason explicitly about connecting components. Considering an architecture then means identifying connectors and components and describing the interplay between them. Since the emphasis is on structure, formalizing an architecture helps in investigating its salient features; formalizations can be proposed on different levels.

The formalization of an architecture permits reasoning about it since it provides precise and abstract models that usually come with analytical techniques. This is in marked contrast to architectural techniques where the shape of an architecture and its architectural parameters are determined experimentally ([29] provides an example for constructing a substantial real life system). Shaw and Garlan [85, Sec.6] discuss architectural formalisms, they distinguish three levels of formalization:

- **The architecture of a specific system.** This permits a precise characterization of the system-level functions that determine the overall product functionality.

- **The formalization of an architectural style.** Through the description of architectural abstractions it becomes possible to analyze various static or dynamic properties of common architectural patterns or reference architectures which are used informally e.g. as reference architectures. Essential ingredients in such a formalization are provided by connectors and by components.

- **A theory of software architecture.** By classifying architectures and representing them with a mathematical machinery, a deductive basis for analyzing systems is provided.

We will focus on the intermediate level and investigate an architecture where the computational elements are represented through relations.

**Relations.** Nondeterministic and probabilistic constructions share, as we have seen in the previous sections, a common structure in representing the Kleisli construction for a monad. The case of non-deterministic relations is covered through the power set functor on the category of sets, and the stochastic case through the functor which assigns
2.3 Case Study: Architectural Modelling Through Monads.

each measurable set the space of all sub-probability measures, as we have seen in the discussion in section 2.2. Thus monads (and their associated Kleisli categories) form the common abstraction for both cases, bringing us into the realm of Moggi’s compelling argumentation [67, 66] that monads form a suitable basis for modelling computations. Consequently, the architectural modelling will be done on the basis of a monad.

**Categories vs. Architectures.** Categories with their emphasis on structure are a suitable formal tool for modelling software architectures. Focussing on structure implies the independence on any representation in a specification or programming language; technically this is achieved through the use of morphisms and functors. Synthesizing a design sometimes means formulating the components and amalgamating them through a suitable colimit, cf. [34]. Wermelinger and Fiadeiro [96] discuss some salient features of an architectural modelling through categories in the context of their modelling mobile programs. Specifically they point out that this approach represents programs as objects, morphisms show how programs can be composed; the explicit use of connectors facilitates the separation of computation and coordination. Moreover they point out that the mechanisms for interconnecting components yielding complex systems are formalized using universal constructs, in this way providing a stage for arguing about these mechanisms formally.

When modelling an architecture, one has to take care at least of the computational components and the connectors. Working in a category, the connectors are represented as objects while the computational components are modelled as morphisms between the objects. Since computations will be represented as monads, the most natural way is representing a component through the work of the corresponding functor $\mathcal{T}$. Here the Kleisli construction enters the game: suppose for simplicity that the input and the output for a component $\lambda$ are modelled respectively through the objects $x$ and $y$, then the computation performed by $\lambda$ is represented through a Kleisli morphism $x \to \mathcal{T}y$. These assignments are described when modelling a particular architecture, and the work of an instance of this architecture is described in terms of these assumptions. We will show how this is done for a pipeline architecture. This architectural style is simple enough to be studied without having to discuss too many technical and architecture specific issues. It is rather general, hence not tied to a particular domain or application, and it is semantically rich enough to illustrate the concepts proposed and investigated here.

**Pipelines.** Filters transform streams of data functionally; each filter has input ports from which data are read, and output ports, to which results are written. Computation is performed incrementally and locally: a portion of the data available at the input ports is read, transformed, and written to the output ports which in turn serve as input ports for other components or as outputs for the system. The filters may be assumed to work concurrently. It is characteristic for this style that the data passing through a filter comes only through its input ports, and leaves only through its output ports, global data are not available. A pipe links an input port to an output port and transmits data from one component to another. Pipelines are in this taxonomy a sub-style which performs the computations without cycles.

Figure 2.1 shows an example for a simple pipeline. The system has two inputs $w_1$ and $w_2$ and two outputs $b_1$ and $b_2$, it has four independent components $1, \ldots, 4$. The edges
are labelled with the types of the inputs the components accept, and produce, resp.: for example, component 1 accepts inputs of type \( X_1 \) and produces outputs of types \( X_3 \) and \( X_4 \), the former serving as an input to component 2 together with an input of type \( X_2 \), the latter serving as an input to component 3 together with an input of type \( X_5 \), which is produced by component 2. The entire system accepts two inputs of type \( X_1 \) and \( X_2 \) and produces two outputs of type \( X_7 \) and \( X_8 \).

We assume that the system forms a directed graph with filters as nodes and pipes as edges. The graph is assumed to be acyclic, so that loops among filters are not permitted, hence we will address what is sometimes called the linear sub-style of the architecture. The common pipelines are usually linear: think of UNIX pipes or of linear arrangements in which data are generated and collected from different sources.

Nevertheless, acyclicity is an assumption which from the mathematical point of view quite notably restricts a general model of pipes and filters. Suppose that we have a component \( C \) which has among its output ports the ports \( r \) and \( s \), \( r \) being a “backward” port, \( s \) being a forward one. Here \( \text{backward} \) means that \( r \)'s output is being fed into the input port of another component the output of which is pipelined directly or indirectly into one of the input ports of component \( C \). With other words: \( C \) lies on a cycle in the graph constituting the system’s topology. The functional character of the components implies that each input is associated with some output for all ports, so that an undefined value at one of these ports must not occur. Modelling this situation requires taking care of this iterative structure, and providing a neutral element of some sorts, indicating that the functional output of \( C \) flows through \( s \) only in certain situations. We will return to this point when discussing possible extensions in section 2.3.7.

**Overview.** The discussion is organized as follows: we make our assumptions on the category we are working in explicit in section 2.3.2, then perform some basic constructions, and relate them to Mac Lane’s monoidal categories, and to Moggi’s strong monads. These constructions are based essentially on the monads representing relations, viz., the Manes, and the Giry monad, resp, for which we did some preparatory work in the sections 2.1 and 2.2. We will base our specific constructions on directed acyclic graphs (\( \text{dags} \)), and we work first on a special class of graphs that we call \( \text{stratified} \). We then demonstrate that assuming stratified graphs is no loss of generality by constructing a
2.3 Case Study: Architectural Modelling Through Monads.

stratified graph from a dag, and by showing that the behavior of the entire system is invariant against stratification. This helps demonstrating that this construction may be used for incrementally and hierarchically constructing pipeline systems in section 2.3.6. After having done all this, we will compare this approach to other people’s work and think a bit about what could be done further along these lines. But before entering the discussion with formal machinery, we will first have a look at a simple example, so that the ideas will become more transparent.

2.3.1 A First Example

Consider the pipeline system of pipes and filters represented through the graph in Figure 2.1 again. It has the roots \( \{w_1, w_2\} \), the leaves \( \{b_1, b_2\} \) and the filters \( \{1, 2, 3, 4\} \). This graph exhibits a little irregularity in that paths to a node from different roots may differ in length. Thus a signal from root \( w_1 \) to node 2 is routed through node 1 before arriving at node 2, whereas a signal originated from root \( w_2 \) arrives directly at node 2. In order to get a uniform treatment, we introduce noops which are to have the effect that all paths from a root to a node have the same length (we will call graphs with this property stratified later on, cf. Definition 2.3.8). Thus the graph is not stratified, but the version in Figure 2.2 is. Note that we have introduced two new artificial nodes \( \Delta_2 \) and \( \Delta_4 \).

We partition the set of nodes into classes \( S_j \) such that a node \( n \) is in \( S_j \) iff the length of a path from a root to \( n \) is exactly \( j \). Hence we have these sets \( S_0, \ldots, S_4 \):

\[
S_0 = \{w_1, w_2\}, S_1 = \{1, \Delta_2\}, S_2 = \{\Delta_4, 2\}, S_3 = \{3, 4\}, S_4 = \{b_1, b_2\}.
\]

We will associate sets with the edges, and relations between the corresponding sets with the nodes. Sets are meant to indicate the type of information flowing along that edge, relations associated with a node to indicate the processing being performed by this node. Hence we have for node 2 relation \( R_2 \) with \( R_2 \subseteq (X_2 \times X_3) \times (X_5 \times X_6) \). The noop nodes are associated with no processing at all, thus the information is just passed unchanged through them, indicated by \( \Delta \), and we have

\[
\Delta_2 := \{\langle x, x \rangle \mid x \in X_2\}, \\
\Delta_4 := \{\langle x, x \rangle \mid x \in X_4\}.
\]
Categorial and Probabilistic Aspects of Stochastic Relations

Work in these partitions proper is characterized by independent processing of the partition’s nodes, thus we take the Cartesian product of the sets labelling the incoming edges as the input sets for the partition, similarly for the output sets. The construction yields e.g. for component $S_2$ this relation:

$$\{ \langle x_2, x_3, x_4, x_5, x_6 \rangle \mid x_4 \in X_4, \langle x_2, x_3, x_5, x_6 \rangle \in R_2 \}$$

The work of the entire pipeline is then described through the product of the relations that represent the work of the individual partitions. This yields a relation which is a subset of $(X_1 \times X_2) \times (X_7 \times X_8)$. The result is, as expected:

$$\{ \langle x_1, x_2, x_7, x_8 \rangle \mid \exists x_3 \in X_3, x_4 \in X_4, x_5 \in X_5, x_6 \in X_6 : 
\langle x_1, x_3, x_4 \rangle \in R_1, \langle x_2, x_3, x_5, x_6 \rangle \in R_2, \langle x_4, x_5, x_7 \rangle \in R_3, \langle x_6, x_8 \rangle \in R_4 \}.$$

This relation may be manipulated further, it may take part in horizontal and vertical operations. Horizontal operations form architectures by concatenating components, so that the resulting pipelines get longer and longer, vertical operations refine an architecture by replacing a component through an entire subsystem (we are a bit in conflict with the notion of horizontal and vertical composition in category theory, cp. [62, II.4, II.5 and XII.3] where these terms are used for the composition of natural transformations for different functors; this terminology is, however, quite graphically used in Software Engineering, too, so we stick to it here). These fundamental architectural operations will be discussed in detail in section 2.3.6.

2.3.2 First Steps

This section serves as a preparation for things to come: we formulate a compatibility condition which relates the product in the category under consideration to the monad which is used for modelling the computations. The category $X^{(n)}$ has as objects $n$-tuples of objects of $X$, and $n$-tuples of morphisms, the composition being defined componentwise. Define functors $\mathcal{G}_T^{(n)}, \mathcal{H}_T^{(n)} : X^{(n)} \to X$ upon setting $(n \geq 1)$

$$\mathcal{G}_T^{(n)}(x_1, \ldots, x_n) := \mathcal{T} x_1 \times \cdots \times \mathcal{T} x_n$$
$$\mathcal{H}_T^{(n)}(x_1, \ldots, x_n) := \mathcal{T} (x_1 \times \cdots \times x_n),$$

and, if $\phi_i : x_i \to y_i$ are morphisms in $X$, then

$$\mathcal{G}_T^{(n)}(\phi_1, \ldots, \phi_n) := \mathcal{T} \phi_1 \times \cdots \times \mathcal{T} \phi_n$$
$$\mathcal{H}_T^{(n)}(\phi_1, \ldots, \phi_n) := \mathcal{T} (\phi_1 \times \cdots \times \phi_n).$$

\(\mathcal{T}\) models the computations performed in the components, and which are partially done in parallel. This in turn will be modelled through finite products. Hence \(\mathcal{T}\) should be naturally related to the product in $X$; the present proposal assumes compatibility which mediates between $\mathcal{T} (x) \times \mathcal{T} (y) \times \mathcal{T} (z)$ and $\mathcal{T} (x \times y \times z)$ using the natural transformation $1_\mathcal{T} : \mathcal{T} \to \mathcal{T}$ and introducing another one between $\mathcal{G}_T^{(2)}$ and $\mathcal{H}_T^{(2)}$. To be specific:
**Definition 2.3.1** Monad $T$ is compatible with the product in $X$ iff there exists a natural transformation $\theta: \mathcal{G}_T^{(2)} \to \mathcal{H}_T^{(2)}$ which makes this diagram commutative:

\[
\begin{align*}
\mathcal{T}(x) \times \mathcal{T}(y) \times \mathcal{T}(z) & \xrightarrow{(\theta \times 1_{\mathcal{T}})(x,y,z)} \mathcal{T}(x) \times \mathcal{T}(y \times z) \\
(\theta \times 1_{\mathcal{T}})(x,y,z) & = \theta(x,y \times z) \\
\mathcal{T}(x \times y) \times \mathcal{T}(z) & \xrightarrow{\theta(x \times y, z)} \mathcal{T}(x \times y \times z)
\end{align*}
\]

$\theta$ is called the mediating transformation.

A mediating transformation $\theta$ spawns a sequence $(\theta^{(n)})_{n \geq 1}$ of natural transformations

\[
\theta^{(n)}: \mathcal{G}_T^{(n)} \to \mathcal{H}_T^{(n)}
\]

in the following way:

\[
\begin{align*}
\theta^{(1)} &: = 1_{\mathcal{T}x} \\
\theta^{(2)} &: = \theta(x,y) \\
\theta^{(n+1)} &: = \theta(x_{1} \cdots x_{n}, x_{n+1}) \circ \left( (\theta^{(n)} \times 1_{\mathcal{T}})_{x_1 \cdots x_{n+1}} \right)
\end{align*}
\]

**Lemma 2.3.2** Suppose that $\mathcal{T}$ is compatible with the product in $X$, and let $\theta$ be the mediating transformation. Defining $\theta$ as above, the sequence $(\theta^{(n)})_{n \in \mathbb{N}}$ has the following properties:

1. $\theta^{(n)}$ is a natural transformation,
2. for all $k, \ell \in \mathbb{N}$ and for all objects $x_i, y_i$ we have

\[
\theta(x_1 \cdots x_k, y_1 \cdots y_\ell) \circ \left( (\theta^{(k)} \times \theta^{(\ell)})_{x_1 \cdots x_k, y_1 \cdots y_\ell} \right) = \theta^{(k+\ell)}_{x_1 \cdots x_k, y_1 \cdots y_\ell}
\]

**Proof**

1. The first part is established by induction on $n$, since the composition of natural transformations is again a natural transformation.
2. The second part is proved by induction on $\ell$, the start of the induction representing just the inductive definition from above. The induction step is established through the commutativity of this diagram:

\[
\begin{align*}
(\mathcal{T}a_1 \times \cdots \times \mathcal{T}a_k) \times (\mathcal{T}b_1 \times \cdots \times \mathcal{T}b_\ell) \times \mathcal{T}(b) & \xrightarrow{\alpha} \mathcal{T}(a_1 \times \cdots \times a_k) \times (\mathcal{T}b_1 \times \cdots \times \mathcal{T}b_\ell) \times \mathcal{T}(b) \\
& \xrightarrow{\beta} \mathcal{T}(a_1 \times \cdots \times a_k \times b_1 \times \cdots \times b_\ell) \times \mathcal{T}(b) \\
& \xrightarrow{\kappa} \mathcal{T}(a_1 \times \cdots \times a_k \times b_1 \times \cdots \times b_\ell \times b) \\
& \xrightarrow{\lambda} \mathcal{T}(a_1 \times \cdots \times a_k \times b_1 \times \cdots \times b_\ell \times b)
\end{align*}
\]
with
\[
\begin{align*}
\alpha & := (\theta^k \times \theta^\ell \times 1_\tau)_{\langle a_1, \ldots, a_k, b_1, \ldots, b_\ell, b \rangle} \\
\beta & := (1_\tau \times \theta)_{\langle a_1 \times \ldots \times a_k, b_1 \times \ldots \times b_\ell \times b \rangle} \\
\gamma & := (\theta^{k+\ell} \times 1_\tau)_{\langle a_1, \ldots, a_k, b_1, \ldots, b_\ell, b \rangle} \\
\delta & := \theta_{\langle a_1 \times \ldots \times a_k \times b_1 \times \ldots \times b_\ell \times b \rangle} \\
\kappa & := (\theta \times 1_\tau)_{\langle a_1 \times \ldots \times a_k, b_1 \times \ldots \times b_\ell \times b \rangle} \\
\lambda & := \theta_{\langle a_1 \times \ldots \times a_k, b_1 \times \ldots \times b_\ell \times b \rangle}
\end{align*}
\]

The upper triangle is commutative because of the induction hypothesis, the lower square is just the condition on \(\theta\) from Definition 2.3.1. Then the assertion follows, since
\[
\lambda \circ \beta \circ \alpha = \delta \circ \kappa \circ \alpha = \delta \circ \gamma,
\]
and
\[
\begin{align*}
\beta \circ \alpha &= (\theta^k \times \theta^{\ell+1})_{\langle a_1, \ldots, a_k, b_1, \ldots, b_\ell, b \rangle} \\
\delta \circ \gamma &= \theta^{k+\ell+1}_{\langle a_1, \ldots, a_k, b_1, \ldots, b_\ell, b \rangle}
\end{align*}
\]

Our sample categories have mediating transformations.

**Lemma 2.3.3**

\[\mathcal{M}(X) \times \mathcal{M}(Y) \ni \langle A, B \rangle \mapsto \{(h_1 h_2, x, y) \mid \langle h_1, x \rangle \in A, \langle h_2, y \rangle \in B\} \in \mathcal{M}(X \times Y)\]

defines a natural transformation that mediates between the functor and the product in Set.

Let us discuss the Giry monad. Define for the measurable subset \(C\) of \(H \times X_1 \times X_2\) and for \(\mu_i \in \mathcal{G}(X_i)\) \((i = 1, 2)\) their \(H\)-product \(\mu_1 \otimes_H \mu_2\) through
\[
(\mu_1 \otimes_H \mu_2)(C) := \int_{H \times X_1} \mu_2(\{\langle h_2, y \rangle \mid \langle h_1 h_2, x, y \rangle \in C\}) \mu_1(d\langle h_1, x \rangle),
\]
then \(\mu_1 \otimes_H \mu_2 \in \mathcal{G}(X_1 \times X_2)\). In fact, we can say more:

**Lemma 2.3.4** The \(H\)-product is associative, it constitutes a natural mediating transformation \(\mathcal{G}^{(2)} \circ \bullet \rightarrow \mathcal{G}^{(2)}\).

**Proof** 1. Associativity of the \(H\)-product is an easy consequence of Fubini’s Theorem on product integration.

2. Let \(f : X \to X'\) and \(g : Y \to Y'\) be measurable maps, then we have for the measures \(\mu_1 \in \mathcal{G}(X), \mu_2 \in \mathcal{G}(Y)\) and for the measurable subset \(C'\) of \(H \times X' \times Y'\)
\[
(\mathcal{G}(f)(\mu_1) \otimes_H \mathcal{G}(g)(\mu_2))(C') = \int_{H \times X} \mu_2(\{\langle h_2, y \rangle \mid \langle h_1 h_2, f(x), g(y) \rangle \in C'\}) \mu_1(d\langle h_1, x \rangle) = (\mu_1 \otimes_H \mu_2)(\{\langle h, x, y \rangle \mid \langle g, f(x), g(y) \rangle \in C'\}) =\]
\[
\mathcal{G}(f \times g)(\langle \langle \mu_1, \mu_2 \rangle \mapsto \mu_1 \otimes_H \mu_2 \rangle)(C').
\]
But this means that the $H$-product is natural for $\mathcal{G}^{(2)}$ and $\mathcal{H}^{(2)}$. It is easy to see that this transformation is mediating. ⊣

The product in $X$ defines together with $T$ an associative operation:

**Definition 2.3.5** Let $\tau : a \to T b$ and $\tau' : a' \to T b'$ be morphisms, then define $\tau \times T \tau' : a \times a' \to T (b \times b')$ upon setting $\tau \times T \tau' := \theta_{(b,b')} \circ \tau \times \tau'$.

Example 2.3.6 will show that $\times T$ does not exhibit the universal properties which would be necessary to form a product (thus the category $X_T$ does not necessarily have finite products, even if $X$ has them).

**Example 2.3.6** Let, for simplicity, $H$ be the trivial monoid $\{1\}$, which we omit from the notation. Suppose that $\times G$ is a product in $\mathcal{M}_G$, and fix two non-empty measurable spaces $X_1$ and $X_2$. There exists a measurable space $X$ and the two projections $p_i : X \to \mathcal{G}(X_i)$ such that, whenever $K_i : S \to \mathcal{G}(X_i) (i = 1, 2)$ is a morphism, we can find a morphism $K : S \to \mathcal{G}(X)$ such that $K_i = p_i * K$ holds for $i = 1, 2$. This means that

$$K_i(s)(B_i) = \int_X p_i(x)(B_i) K(s)(dx)$$

always holds. Now let always $K_i(s)(X_i)$ equal 1. This implies that $p_i(x)(X_i) = 1$ is true $K(s)$-almost everywhere for each $s$, and for each $K$ which can be so constructed. Note that $p_1, p_2$ do not depend on the specific choice of $K_1, K_2$. But then we have for any $L_i : S \to \mathcal{G}(X_i)$ with product $L$:

$$L_1(s)(X_1) = \int_X p_1(x)(X_1) L(s)(dx) = \int_X p_2(x)(X_2) L(s)(dx) = L_2(s)(X_2).$$

Since we cannot always maintain $L_1(s)(X_1) = L_2(s)(X_2)$ it follows that $\mathcal{M}_\mathcal{G}$ does not have finite products. ⊤

We have, however:

**Corollary 2.3.7** Let $\tau : a \to T b, \tau : a' \to T b', \tau : a'' \to T b''$ be morphisms in $X$, then

$$(\tau \times T \tau') \times T \tau'' = \tau \times T (\tau' \times T \tau'').$$

**Proof** Lemma 2.3.2 shows that both sides of this equation equal

$$\theta_{(b,b',b'')}^{(3)} \circ (\tau \times \tau' \times \tau'').$$

There are other constructions in the literature that relate the product in a category with a monad. Mac Lane [62, Ch. XI.2] defines a *monoidal functor* between monoidal categories which comes close to the compatibility definition proposed here for an endofunctor, where the rôle of the tensor product there is played by the product here. The present definition does not require any conditions on the terminal elements and its image under the functor, thus it is weaker. Mac Lane formulates a transformation quite similar to the...
Categorial and Probabilistic Aspects of Stochastic Relations

one given in Lemma 2.3.2. The proof in [62] refers, however, to a coherence theorem and appears a bit inaccessible by not making the construction transparent. Consequently, a direct proof is given here. Moggi [67, Def. 3.2] on the other hand defines a strong monad in a category which is closed under finite products by postulating the existence of a natural transformation \( t_{a,b} : a \times \mathcal{T}b \rightarrow \mathcal{T}(a \times b) \) having some properties which relate \( \mathcal{T} \) to the product in the category (it called a tensorial strength). In [66, 3.2.3] it is shown how the tensorial strength induces a natural transformation \( \gamma \cdot \eta^{(2)} \) in the terminology used here. Barbosa [8, 3.51 – 3.55] discusses a strength catalogue for distributive categories.

2.3.3 The Basic Construction

We will associate a computation to a pipeline by composing computations performed in its components. This construction will be first carried out due to technical reasons for graphs that exhibit a certain regularity: the nodes are partitioned into layers so that the information flows strictly from one layer to the next one. This restriction is introduced for reasons of synchronization: the inputs at each port of a component in a layer are uniformly available at the same time, so are the outputs. It makes modelling somewhat easier, but it is really only a technical device. We remove it in section 2.3.5, after we have shown in section 2.3.4 how to manipulate a dag so that it is satisfied. Fix in this section a finite dag \( G = (V,E) \) with roots \( W \) and leaves \( B \); for convenience we assume the set \( V \) of nodes to be somehow linearly ordered. Put for node \( n \)

\[
\bullet n := \{m \in V \mid \langle n, m \rangle \in E\}
\]

\[
\bullet n := \{m \in V \mid \langle m, n \rangle \in E\}
\]

as the sets of nodes which have an edge into that node or out of it, resp. \( G \) is not supposed to have any isolated nodes, i.e., nodes \( n \) with \( \bullet n \cup \bullet n = \emptyset \). We define sets \((S_j)_{0 \leq j \leq k}\) through

\[
S_0 := W,
\]

\[
S_{j+1} := \{n \in V \mid \bullet n \subseteq S_j \} \quad (j \geq 0).
\]

Definition 2.3.8 The dag \( G = (V,E) \) is called stratified iff the sets \((S_j)_{0 \leq j \leq k}\) form a partition of \( V \) for some \( k \). The maximal index \( k \) such that \( S_k \neq \emptyset \) is denoted by \( \Lambda(G) \).

Let for the rest of this section \( G \) be a stratified graph.

Lemma 2.3.9 The set of inputs \( \{\langle m, n \rangle \mid \langle m, n \rangle \in E, n \in S_j\} \) into the set \( S_j \) of nodes equals the set of outputs \( \{\langle m, \ell \rangle \mid \langle m, \ell \rangle \in E, m \in S_{j-1}\} \) from the sets \( S_{j-1} \) for \( j \geq 1 \).

Proof This follows directly from the fact that \( G \) is stratified. ⊠

This observation shows that each node \( n \) is in some uniquely determined set \( S_j \). If \( n \) is an inner node (thus if \( j > 0 \) and \( j < k \)), then \( \langle m, n \rangle \in E \) implies \( m \in S_{j-1} \), and \( \langle n, m \rangle \in E \) implies \( m \in S_{j+1} \). Depicting \( S_0, \ldots, S_k \) as blocks from left to right, information flows into \( n \) only from nodes in \( S_{j-1} \), thus from nodes on the left, and flows from \( n \) only into nodes in \( S_{j+1} \), hence into nodes on the right. We associate now objects from category \( X \) with edges, and nodes with morphisms in \( X_{\mathcal{T}} \). To be specific, each edge \( \langle k, n \rangle \in E \) is assigned an object \( \gamma_{\langle k, n \rangle} \) in \( X \). If \( \mathcal{T} \) is the Manes
functor $\mathcal{M}$, this means that an edge $(k, n)$ is assigned a set which represents the flow from node $k$ to node $n$. For $\mathcal{T}$ as the Giry functor $\mathcal{G}$, the edge is assigned a measurable space which also represents the flow along this edge: if it is used as an input, then it is the sample space of all inputs for a probabilistic relation, if it is used as an output, then it represents the space of all probability measures over this space, cf. Example 2.3.10.

The input to node $n$ and the output from this node are then reflected respectively through the respective products

$$i(\gamma, n) := \prod_{k \in \bullet n} (\gamma(k, n) \mid k \in n \notin W)$$

$$o(\gamma, n) := \prod_{k \in n \bullet} (\gamma(n, k) \mid k \in n \notin B)$$

Each inner node $n$ is labelled with a Kleisli morphism $a(\gamma, n) : i(\gamma, n) \to \mathcal{T}(o(\gamma, n))$, so that $a(\gamma, n)$ models the work being performed by node $n$.

**Example 2.3.10** Suppose that $\bullet n = \{m_1, \ldots, m_r\}$ and $n \bullet = \{\ell_1, \ldots, \ell_s\}$.

1. For the Manes monad we assign sets $X_1, \ldots, X_r$ to the edges $(m_1, n), \ldots, (m_r, n)$ and sets $Y_1, \ldots, Y_s$ to the edges $(n, \ell_1), \ldots, (n, \ell_s)$. The node $n$ itself is assigned a relation $a(\gamma, n) \subseteq (X_1 \times \cdots \times X_r) \times (H \times Y_1 \times \cdots \times Y_s)$.

Suppose that $(x_1, \ldots, x_r) \in X_1 \times \cdots \times X_r$ is an input to node $n$ which is related to output $(h, y_1, \ldots, y_s)$. That tuple represents the node’s work, and we have two kinds of results: the tuple $(y_1, \ldots, y_s)$ which in turn is being communicated to other nodes in the pipeline, and $h \in H$ which may be interpreted as immediate result which could be read off this processing element. Proposition 2.1.1 indicates that all these results, which will not be communicated as input to other filters, will be accumulated as control percolates through the system.

2. For the Giry monad, $X_i$ and $Y_j$ are measurable spaces, and $a(\gamma, n) : X_1 \times \cdots \times X_r \leadsto H \times Y_1 \times \cdots \times Y_s$

is a stochastic relation. Thus for an input $(x_1, \ldots, x_r) \in X_1 \times \cdots \times X_r$, and for a measurable $B \subseteq H \times Y_1 \times \cdots \times Y_s$ we get $a(\gamma, n)(x_1, \ldots, x_r)(B)$ as the probability that the computation in node $n$ terminates, and that $(h, y_1, \ldots, y_s)$ will be a member of $B$; the interpretation of the components for this tuple is the same as above.

This indicates that a relational environment for modelling the basic scenario for a pipeline architecture is provided, capturing both the nondeterministic and the probabilistic case. ◊

**Definition 2.3.11** Call $(\mathcal{G}, \gamma)$ a pipeline system (abbreviated as PF-system) over the monad $(\mathcal{T}, \epsilon, m)$ iff the following conditions hold:

- $\mathcal{G} = (V, E)$ is a directed graph with $W$ and $B$ as the sets of roots, and leaves, resp.
- $\forall(n, m) \in E : \gamma(n, m)$ is an object in $X$,
- $\forall n \in V \setminus (W \cup B) : a(\gamma, n) : i(\gamma, n) \to \mathcal{T}(o(\gamma, n))$ is a morphism in $X$
The system \( (G, \gamma) \) is called stratified iff \( G \) is stratified.

Since the monad will be fixed in the sequel, we will not mention it explicitly when talking about PF-systems; unless explicitly mentioned, PF-systems will be stratified in this section.

Now define for \( 0 < j \leq k \) the object

\[
g(\gamma, S_j) := \prod \{ i(\gamma, n) \mid n \in S_j \},
\]

then \( g(\gamma, S_j) \) indicates the kind of flow into \( S_j \) (which is, because of Observation 2.3.9, the flow out of \( S_{j-1} \)); hence the component \( S_j \) has what could be called the input signature \( g(\gamma, S_{j-1}) \) and the output signature \( g(\gamma, S_j) \).

The work being done in \( S_j \) can be represented through the Kleisli morphism

\[
A(\gamma, S_j) : g(\gamma, S_{j-1}) \to \mathbb{T}(g(\gamma, S_j)),
\]

with

\[
A(\gamma, S_j) := \theta^{(#S_j)} \circ \prod \{ a(\gamma, n) \mid n \in S_j \},
\]

where \( \theta \) is the natural transformation which mediates between \( \mathbb{T} \) and the product in \( X \), cf. Definition 2.3.1. The linear order on \( V \) makes \( \theta^{(#S_j)} \) uniquely determined. The work of the entire system is then represented through

\[
P(G, \gamma) := A(\gamma, S_{k-1}) \ast \ldots \ast A(\gamma, S_1).
\]

The construction shows that

\[
P(G, \gamma) : g(\gamma, S_1) \to \mathbb{T}(g(\gamma, S_k))
\]

is a Kleisli morphism between the inputs to the system and the outputs from it, thus represents the system's work.

The example that follows discusses a particular pipeline system, stratifies the graph and exercises the construction proposed here for the Giry monad. The set theoretic case has already been dealt with in section 2.3.1.

**Example 2.3.12** We discuss the example outlined in section 2.3.1 (Figure 2.2) again, this time for the stochastic case. We assume that the monoid carries a measurable structure which makes multiplication measurable, cf. section 2.2.1. The \( \Delta_i \) \((i = 2, 4)\) are Dirac kernels: we put

\[
\Delta_i(x) := \delta_{(1,x)} \in \mathbb{G}(X_i).
\]

Node \( n \) is this time represented through a stochastic relation \( K_n \) between the appropriate sets, e.g.,

\[
K_2 : X_2 \times X_3 \leadsto H \times X_5 \times X_6.
\]

The construction gives then e.g. for component \( S_2 \):

\[
A(\gamma, S_2)(x_2, x_3, x_4) = K_2(x_2, x_3) \otimes_H \Delta_4(x_4),
\]
thus the probability that the computation in components $S_2$ will give an element of the measurable set $D \subseteq H \times X_4 \times X_5 \times X_6$ after input of $(x_1, x_2, x_3) \in X_1 \times X_2 \times X_3$ is computed as

$$A(\gamma, S_2)(x_2, x_3, x_4)(D) = \int_{H \times X_5 \times X_6} \Delta_4(x_4) \left( \{ \langle h_4, x'_4 \rangle \mid \langle h_2h_4, x'_4, x_5, x_6 \rangle \in D \} \right) K_2(x_2, x_3)(d\langle h_2, x_5, x_6 \rangle) = K_2(x_2, x_3)(\{ \langle h_2, x_5, x_6 \rangle \mid \langle h_2, x_4, x_5, x_6 \rangle \in D \}).$$

Let $f : H \times X_4 \times X_5 \times X_6 \to \mathbb{R}$ be a bounded and measurable function, then a computation of the Kleisli product according to Proposition 2.2.6 shows that $(x_1 \in X_1, x_2 \in X_2)$

$$\int_{H \times X_4 \times X_5 \times X_6} f d((A(\gamma, S_2) * A(\gamma, S_1)))(x_1, x_2)
= \int_{H \times X_3 \times X_4} \int_{H \times X_5 \times X_6} f(gh, x_4, x_5, x_6) K_2(x_2, x_3)(d\langle h, x_5, x_6 \rangle) K_1(x_1)(d\langle g, x_3, x_4 \rangle).$$

We get in this way for $(x_1, x_2) \in X_1 \times X_2$ and the measurable subset $F \subseteq H \times X_7 \times X_8$

$$P(G, \gamma)(x_1, x_2)(F)
= \int_{H \times X_4 \times X_5 \times X_6} A(\gamma, S_3)(x_4, x_5, x_6)(\{ \langle h, x_7, x_8 \rangle \mid \langle gh, x_7, x_8 \rangle \in F \}) \times d(A(\gamma, S_2) * A(\gamma, S_1))(x_1, x_2)(d\langle g, x_4, x_5, x_6 \rangle)
= \int_{H \times X_3 \times X_4} \int_{H \times X_5 \times X_6} \int_{H \times X_7} K_4(x_6)(\{ \langle h_1, x_8 \rangle \mid \langle g_1gh_1, x_7, x_8 \rangle \in F \}) \times K_3(x_4, x_5)(d\langle g, x_7 \rangle) K_2(x_2, x_3)(d\langle h, x_5, x_6 \rangle) K_1(x_1)(d\langle g_1, x_3, x_4 \rangle)$$

as the work of the entire pipeline. ◊

We will now prepare for removing the condition that a PF-system should be stratified.

### 2.3.4 Stratifying Graphs

The assumption in carrying out the basic construction in section 2.3.3 has been that the graph underlying the PF-system is stratified. But graphs rarely are, so it becomes necessary to make provisions for generalizing the construction to general directed graphs. The strategy is to devise a way of stratifying a graph, to perform the construction on the new graph, and to make sure that all graphs that are stratified versions of the given one perform the same work. The present section is auxiliary in character and provides an algorithm for stratifying, section 2.3.5 will do the generalization.

Algorithm 2.3.13 produces from $G = (V, E)$ a stratified graph $G' = (V', E')$ with $V \subseteq V'$ and $E' \cap (V \times V) \subseteq E$. It assumes that $G$ does not have any isolated nodes, and that each node lies on a path from a root to a leaf. We assume that we have a source $Q$ of fresh nodes which is disjoint from $V \cup E$; invoking the function newQ() will produce a fresh node. The map $\alpha$ initially gives the in-degree of a node; we use some auxiliary values which will be needed and discussed in the sequel.
Algorithm 2.3.13
Edges := E; Nodes := V; zeta := 0;
while Edges ≠ ∅ do
    forall n ∈ range Edges do
        H(n) := {a | ⟨a,n⟩ ∈ Edges, α(a) = 0};
    od;
    forall a ∈ domain H do
        d(a) := zeta;
    od;
    Edges := Edges \ {⟨a,n⟩ | ⟨a,n⟩ ∈ Edges, α(a) = 0};
    forall n ∈ Nodes do
        r := # {k | ⟨k,n⟩ ∈ Edges};
        if r = 0 then
            α(n) := 0;
        else
            for j := r + 1 to α(n) do
                choose m from H(n);
                H(n) := H(n) \ {m};
                q := newq();
                α(q) := 0; V := V ∪ {q};
                E := (E \ {⟨m,n⟩}) ∪ {⟨m,q⟩, ⟨q,n⟩};
                Edges := Edges ∪ {⟨q,n⟩};
            od; -- forall
        fi;
    od; -- forall
    zeta := zeta + 1;
od; -- while ♠

Figure 2.3: Algorithm Stratify

Thus we iterate over all edges, removing roots as we go; for a node n we use H(n) for recording which nodes have edges leading into n that will be removed. When we see that a node n has no longer any edges having n as a target, this node will be promoted to a root (and removed in due course); promotion to a root means changing the in-degree α(n) to 0. If it turns out, however, that there are still edges going into that node (note that in this case #H(n) equals α(n) − r), we replace each edge ⟨m,n⟩ by a pair of edges ⟨m,q⟩ and ⟨q,n⟩, where q is a fresh node which is put into the set V of nodes.

Since each dag has roots, and since G is assumed to have no isolated nodes, it is not difficult to see that Algorithm 2.3.13 terminates. It is also evident that the new graph G′ = (V′, E′) has the given one as a subgraph in the sense that V ⊆ V′ and E′ ∩ (V × V) ⊆ E both hold.

Lemma 2.3.14 Let n ∈ E′ be a node in G′, and assume that there exists a path from a root of G′ to n. Then this path has length d(n).

Proof 1. We proceed by induction on the value of zeta. The begin is trivial, since exactly the roots of G are removed, and no new roots are introduced.
2. Now let \( \zeta = k \), and assume that \( d(n) \) equals \( k + 1 \). This means that \( \alpha(n) \) is set to 0 when \( \zeta \) has the value \( k \). We distinguish the cases that \( n \) is a new node introduced in this step from the case that \( \alpha(n) \) is set to 0 because \( r = 0 \) holds.

- If \( n \) is a new node, we can find an edge \( \langle m_1, n_1 \rangle \) which gave rise to this creation, hence that edge is replaced by the pair of edges \( \langle m_1, n \rangle \) and \( \langle n, n_1 \rangle \). Edge \( \langle m_1, n_1 \rangle \) is a member of the set \( \text{Edges} \) before control enters the body of the actual loop, thus will be removed. The induction hypothesis makes sure that each path from a root to \( m \) in the graph constructed so far has length \( k \), thus \( d(n) = k + 1 \).

- If \( n \) is no new node, the assumption that there is a path in the new graph to \( n \) implies that, since there is no node \( m \) with \( \langle m, n \rangle \in \text{Edges} \), there are edges \( \langle m, n \rangle \) which have been deleted in the step before. For all these \( m \) we have \( d(m) = k \). In the new graph have all these nodes \( m \) the property that each path from a root to them has length \( k \).

This implies the assertion. \( \dashv \)

An immediate consequence of Observation 2.3.14 is

**Proposition 2.3.15** Algorithm 2.3.13 produces a stratified graph.

**Proof** Using the notation from above, put \( S_j := \{ n \in V' \mid d(n) = j \} \). Then \( S_0 \) is the set of roots for \( \mathcal{G}' \) as well as for \( \mathcal{G} \), and if node \( n \) is in \( S_{j+1} \), then all its predecessors (w. r. t. \( \mathcal{G}' \)) are in \( S_j \). These sets are mutually disjoint, and \( S_{k'} = \emptyset \) for all \( k' \geq k \) for some minimal index \( k \). Since \( n \in S_{d(n)} \) holds for each node \( n \in V' \), we see that \( (S_j)_{0 \leq j \leq k} \) forms a partition of \( V' \). \( \dashv \)

Armed with this tool, we now enter the discussion of the general case.

### 2.3.5 The General Case

We will demonstrate that all the stratified PF-systems which can be constructed from a given one will do the same work, so that this morphism is an invariant, and that it is sensible to assign it to a non-stratified PF-system as its work. This has as a remarkable consequence that two constructions can be carried out that help in composing larger systems from smaller ones: we show in section 2.3.6 how two PF-systems can be glued together (as a horizontal extension), and that hierarchical refinement is available as construction technique, permitting the expansion of a node by an entire subsystem. This is a vertical extension.

Both the PF-system \( \langle \mathcal{G}, \gamma \rangle \), and \( \mathcal{G} = \langle V, E \rangle \) as the graph underlying it are fixed. The sets \( W \) and \( B \) denote the roots, and the leaves of \( \mathcal{G} \), resp. We fix also the set \( Q \) which serves as a reservoir of fresh nodes for stratification.

We begin with an adaptation of Algorithm 2.3.13 to PF-systems by taking the labels for edge and nodes coming with such a system into account. To be specific, suppose we replace an edge \( \langle m, n \rangle \) from the set of edges by the pair \( \langle m, q \rangle \) and \( \langle q, n \rangle \) with the fresh node \( q \in Q \). Then we put

\[
\gamma(m,q) := \gamma(m,n);
\gamma(q,n) := \gamma(m,n);
\alpha(\gamma, q) := e_{\gamma(q,n)}.
\]
Categorial and Probabilistic Aspects of Stochastic Relations

(remember that \(e\) denotes the unit for the corresponding monad). Thus if the edge \(\langle m, n \rangle\) carries type \(a\), where \(a\) is an object in \(X\), then the new edges carry this type, and the node inserted is assigned the Kleisli morphism \(e_a\); note that the natural transformation \(e\) provides the identities in the Kleisli category \(X_T\). In terms of pipelines, by inserting \(e_{\gamma\langle q, n \rangle}\) we insert a noop into the system, since the filter introduced in this way evidently does not do any other work than transporting inputs unchanged to outputs. In this way we obtain from \(\langle G, \gamma \rangle\) a stratified PF-system \(\langle G_1, \gamma \rangle\), reusing \(\gamma\) for simplicity.

The graph constructed by Algorithm 2.3.13 is an extension of the given graph. This is made precise now.

**Definition 2.3.16** The graph \(G' = (V', E')\) is called a \(Q\)-extension to \(G\) iff

1. \(G'\) is stratified with \(E' \cap (V \times V) \subseteq E\), and \(G'\) has the same roots as \(G\),
2. \(V \subseteq V'\), and \(V' \setminus V \subseteq Q\),
3. if \(\langle n, m \rangle \in E \setminus E'\), then there exists a unique path \(n = q_0, \ldots, q_k = m\) from \(n\) to \(m\) in \(G'\) with \(\langle q_i, q_{i+1} \rangle \in E'\) for \(0 \leq i < k\),
4. for all \(q \in V' \setminus V\), \(#(\bullet q) = #(q \bullet) = 1\).

Thus a \(Q\)-extension has new nodes from the fountain \(Q\) of nodes only, an edge in \(E\) is either an edge in \(E'\), or its endpoints are connected through a unique path that runs entirely through \(Q\) (apart from the endpoints, of course). The new nodes in \(G'\) do not have a rich social life by being neighbor to only two other nodes, thus such a node receives inputs from exactly one node and propagates it to a unique other node.

**Lemma 2.3.17** The graph constructed from Algorithm 2.3.13 is an \(Q\)-extension to \(G\).

**Proof** If \(G'\) is the graph constructed from \(G\), then \(G'\) has been shown to be stratified in Proposition 2.3.15. The construction makes sure that the other conditions from Definition 2.3.16 are satisfied. \(\dashv\)

Any \(Q\)-extension can be decorated as indicated above: the nodes from \(Q\) receive \(e_x\) as their function, where \(x\) is an appropriate object which labels the edges leading into that node, and out of it, resp. This leads to the notion of an \(Q\)-extension to a PF-system which will not be formally defined since the definition is obvious (the reader is invited to formulate it).

We want to establish that the work of a PF-system is an invariant for all \(Q\)-extensions to a given PF-system. For this we should make sure that the composition of Kleisli morphisms and the operation \(\times_T\) which resembles a product so closely relate to each other like composition and product:

**Definition 2.3.18** The monad \(\langle T, e, m \rangle\) satisfies the \(\sharp\)-condition iff

1. \(e_{a \times b} = e_a \times_T e_b\) for all objects \(a\) and \(b\) in \(X\),
2. for the morphisms \(f_i : a_i \to T b_i\), \(g_i : b_i \to T c_i\) \((i = 1, 2)\) the equality

\[
(g_1 \times_T g_2) * (f_1 \times_T f_2) = (g_1 * f_1) \times_T (g_2 * f_2)
\]

holds.
Thus the identity on \( a \times b \) in \( X_T \) is obtained from the respective identities on \( a \) and \( b \) by performing the \( \times_T \)-operation. In terms of computation, combining the identities on \( a \) and on \( b \) independently to a component yields the identity in \( a \times b \). The second condition explains the name: viewing the Kleisli composition \(*\) as a horizontal operation along the flow of information which indicates piping, and \( \times_T \) as a vertical operation modelling independent composition, the equation is visualized in Figure 2.4. Hence piping of composed computations is tantamount to composing piped computations.

Let us investigate our reference categories:

**Proposition 2.3.19** Both the Manes and the Giry category satisfy the \( \#\)-condition, provided the monoid \( H \) which comes with the respective monads is commutative.

**Proof**

1. The first condition is readily established for both monads.
2. Let \( R_i : A_i \to \mathfrak{M}(B_i), S_i : B_i \to \mathfrak{M}(C_i) (i = 1, 2) \) be morphisms with \( \mathfrak{M} \) as the functor underlying the Manes monad. Then these equalities hold for \( (a_1, a_2) \in A_1 \times A_2: \)

\[
((S_1 \times \mathfrak{M} S_2) * (R_1 \times \mathfrak{M} R_2))(a_1, a_2) = \\
\{ (h_1 h_2 h_3 h_4, c_1, c_2) \mid \exists b_1, b_2 : (h_1, b_1) \in R_1(a_1), (h_2, b_2) \in R_2(a_2), \\
\quad (h_3, c_1) \in S_1(b_1), (h_4, c_2) \in S_2(b_2) \},
\]

and

\[
((S_1 * R_1) \times \mathfrak{M} (S_2 * R_2))(a_1, a_2) = \\
\{ (h_1 h_2 h_3 h_4, c_1, c_2) \mid \exists b_1, b_2 : (h_1, b_1) \in R_1(a_1), (h_2, c_2) \in S_2(b_2), \\
\quad (h_3, b_2) \in R_2(a_2), (h_4, c_2) \in S_2(b_2) \}.
\]

3. Let \( K_i : A_i \to \mathfrak{G}(B_i), L_i : B_i \to \mathfrak{G}(C_i) (i = 1, 2) \) be morphisms with \( \mathfrak{G} \) as the functor underlying the Manes monad. Here \( A_i, B_i, C_i \) are measurable spaces, the monoid \( H \) is assumed to be measurable as well. Now

\[
((L_1 \times \mathfrak{G} L_2) * (K_1 \times \mathfrak{G} K_2))(a_1, a_2)
\]
Lemma 2.3.20

is a finite measure on $H \times C_1 \times C_2$, and so is

$$((L_1 \cdot K_1) \times_{\mathfrak{M}} (L_2 \cdot K_2))(a_1, a_2),$$

hence it is sufficient for establishing equality to show that the integrals for an arbitrary measurable and bounded function $\psi : H \times C_1 \times C_2 \to \mathbb{R}$ coincide. A calculation using Fubini’s Theorem on product integration establishes that

$$\int_{H \times C_1 \times C_2} \psi \, d \left( (L_1 \cdot \emptyset)(K_1 \cdot \emptyset) \right)(a_1, a_2) =$$

$$\int_{H \times B_1} \int_{H \times B_2} \int_{H \times C_1} \int_{H \times C_2} \psi(h_1 h_2 h_3 h_4, c_1, c_2) L_2(b_2)(d<h_4, c_2) L_1(b_1)(d<h_3, c_1) \times K_2(a_2)(d<h_2, b_2) K_1(a_1)(d<h_1, b_1)$$

and

$$\int_{H \times C_1 \times C_2} \psi \, d \left( (L_1 \cdot \emptyset)(K_1 \cdot \emptyset) \right)(a_1, a_2) =$$

$$\int_{H \times B_1} \int_{H \times C_1} \int_{H \times B_2} \int_{H \times C_2} \psi(h_1 h_2 h_3 h_4, c_1, c_2) K_2(a_2)(d<h_3, b_2) \times L_1(b_2)(d<h_2, c_1) K_1(a_1)(d<h_1, b_1).$$

4. These equalities establish the claim. It is interesting to observe in which way in both cases the roles of $h_2$ and $h_3$ get interchanged, reflecting the way in which morphisms change positions. \(\dashv\)

An easy induction using the second assertion in Lemma 2.3.2 establishes that the Kleisli identity on $a_1 \times \cdots \times a_n$ can be calculated through the identities on the components. The $\sharp$-condition makes also sure that we may shift computations between products (the easy inductive proof is left to the reader):

**Lemma 2.3.20** Assume that the $\sharp$-condition holds. Then

1. The equality

$$\epsilon_{a_1 \times \cdots \times a_n} = \epsilon_{a_1} \times \tau \cdots \times \tau \epsilon_{a_n}$$

holds for all objects $a_1, \ldots, a_n$ in $X$,

2. If $\sigma_i : a_i \to \Sigma b_i$ and $\tau_i : b_i \to \Sigma c_i$ are morphisms in $X$, then

$$(\tau_1 \times \tau_2 \cdots \times \tau_n) \ast (\sigma_1 \times \sigma_2 \cdots \times \sigma_n) =$$

$$(\tau_1 \times \sigma_1 \times \tau_2 \cdots \times \sigma_j \cdot \tau_j \ast \sigma_j) \times \tau_{j+1} \times \cdots \times \tau_n) \ast$$

$$(\sigma_1 \times \tau_1 \times \sigma_2 \cdots \times \sigma_j \ast \tau_j \ast \sigma_j) \times \sigma_{j+1} \times \cdots \times \sigma_n)$$

$$(\tau_1 \times \sigma_1 \times \tau_2 \cdots \times \sigma_j \ast \tau_j \ast \sigma_j) \times \sigma_{j+1} \times \cdots \times \sigma_n) \ast$$

$$(\sigma_1 \times \tau_1 \times \sigma_2 \cdots \times \tau_j \ast \sigma_j) \times \sigma_{j+1} \times \cdots \times \sigma_n).$$

\(\dashv\)
The equations in part 2 of Lemma 2.3.20 are useful in our context: \( \sigma_1 \times_\mathcal{T} \ldots \times_\mathcal{T} \sigma_n \) and \( \tau_1 \times_\mathcal{T} \ldots \times_\mathcal{T} \tau_n \) represent the computations in consecutive blocks of a PF-system. Then we may shift the computation of a component out of a block into the next or the previous one without changing the result; shifting means among others replacing the morphism by the appropriate identity. We will use this observation in the proof of Proposition 2.3.23 for establishing the invariance result.

From now on we assume that the \( \# \)-condition is satisfied.

**A Little Digression.** In fact, we can say more about representing \( \times_\mathcal{T} \)-products of morphisms: they can be written as Kleisli-products of a very special kind. The discerning reader will no doubt observe that the kind of representation derived from the discussion that follows will not be needed for the present constructions of PF-systems. It appears to be interesting, nevertheless.

**Definition 2.3.21** Assume \( n > 1 \), let \( \tau_i : a_i \to \mathcal{T}b_i \) be morphisms for \( 1 \leq i \leq n \) and let \( \xi \) be a permutation of \( \{1, \ldots, n\} \). Then \( \langle \sigma_1, \ldots, \sigma_n \rangle \) is the \( \xi \)-expansion of \( \langle \tau_1, \ldots, \tau_n \rangle \) iff \( \sigma_j \) can be written as \( \zeta_{j,1} \times_\mathcal{T} \ldots \times_\mathcal{T} \zeta_{j,n} \) such that

1. each \( \zeta_{j,i} \) is either \( e_{a_i}, e_{b_i} \) or one of \( \tau_1, \ldots, \tau_n \),
2. \( \zeta_{j,k} \in \{\tau_1, \ldots, \tau_n\} \) iff \( \xi(j) = k \),
3. if \( \zeta_{j,k} = \tau_i \), then

\[
\zeta_{\ell,k} = \begin{cases} 
  e_{a_i}, & \ell > j \\
  e_{b_i}, & \ell < j.
\end{cases}
\]

For example, the permutation \((13)(2)\) of \( \{1, 2, 3\} \) corresponds to

\[
\begin{pmatrix}
  \zeta_{1,1} & \zeta_{1,2} & \zeta_{1,3} \\
  \zeta_{2,1} & \zeta_{2,2} & \zeta_{2,3} \\
  \zeta_{3,1} & \zeta_{3,2} & \zeta_{3,3}
\end{pmatrix} =
\begin{pmatrix}
  e_{b_1} & e_{b_2} & \tau_3 \\
  e_{b_1} & \tau_2 & e_{a_3} \\
  \tau_1 & e_{a_2} & e_{a_3}
\end{pmatrix}
\]

Thus if \( \langle \sigma_1, \ldots, \sigma_n \rangle \) is a \( \xi \)-expansion of \( \langle \tau_1, \ldots, \tau_n \rangle \), then assuming \( \xi(j) = i \), \( \sigma_j \) can be written as

\[ e_{b_1} \times_\mathcal{T} \ldots \times_\mathcal{T} e_{b_{i-1}} \times_\mathcal{T} \tau_i \times_\mathcal{T} e_{a_{i+1}} \times_\mathcal{T} \ldots \times_\mathcal{T} e_{a_n} \]

indicating that \( \tau_i \) is doing its work, whereas \( \tau_1, \ldots, \tau_{i-1} \) did do their work already (thus the identity on the range is incorporated) and that \( \tau_{i+1}, \ldots, \tau_n \) will still have to do their work (hence the identity of the respective domains are incorporated into the \( \times_\mathcal{T} \)-product).

**Lemma 2.3.22** Let, under the assumptions of Definition 2.3.21, \( \langle \sigma_1, \ldots, \sigma_n \rangle \) be an \( \xi \)-expansion of \( \langle \tau_1, \ldots, \tau_n \rangle \), then

\[ \tau_1 \times_\mathcal{T} \ldots \times_\mathcal{T} \tau_n = \sigma_1 \ast \ldots \ast \sigma_n \]

holds.
Proof 1. The proof proceeds by induction on \( n \). For \( n = 2 \) the only \( \xi \)-expansions of \( \langle \tau_1, \tau_2 \rangle \) are \( \langle e_{b_1} \times \tau_2, \tau_1 \times e_{a_2} \rangle \) and \( \langle \tau_1 \times \tau e_{b_2} e_{a_1} \times \tau \tau_2 \rangle \). The \( \sharp \)-conditions then permits directly establishing the claim.

2. The inductive step considers the \( \xi \)-expansion \( \langle \sigma_1, \ldots, \sigma_{n+1} \rangle \) of \( \langle \tau_1, \ldots, \tau_{n+1} \rangle \). Then \( \xi(j_{n+1}) = n + 1 \), and we can write

\[
\sigma_i = \begin{cases} 
\sigma'_i \times e_{b_{n+1}}, & i < j_{n+1} \\
\sigma'_i \times e_{a_{n+1}}, & i > j_{n+1}
\end{cases}
\]

for some \( \sigma'_i \). It is easy to see that

\[
\langle \sigma'_1, \ldots, \sigma'_{j_{n+1}-1}, \sigma'_{j_{n+1}+1}, \ldots, \sigma_{n+1} \rangle
\]

is an \( \xi' \)-expansion for \( \langle \tau_1, \ldots, \tau_n \rangle \), where \( \xi' \) is the permutation of \( \{1, \ldots, n\} \) derived from \( \xi \). Now write \( \sigma_{j_{n+1}} = e_{c_1} \times \tau \ldots \times \tau e_{c_n} \) where \( c_i \in \{a_i, b_i\} \) is suitably chosen according to the definition of the expansion, then we have by the induction hypothesis, by the \( \sharp \)-condition, and by Lemma 2.3.20

\[
\sigma_1 \ast \ldots \ast \sigma_{n+1} = (\sigma'_1 \times e_{b_{n+1}}) \ast \ldots \ast (\sigma'_{j_{n+1}-1} \times e_{b_{n+1}}) \ast \\
\ast \sigma_{j_{n+1}} \ast (\sigma'_{j_{n+1}+1} \times e_{a_{n+1}}) \ast \ldots \ast (\sigma'_{n+1} \times e_{a_{n+1}})
\]

\[
= (\sigma'_1 \ast \ldots \ast \sigma'_{j_{n+1}-1} \ast (e_{c_1} \times \tau \ldots \times \tau e_{c_n}) \times \tau \tau_{n+1} \ast \\
\ast (\sigma'_{j_{n+1}+1} \ast \ldots \ast \sigma'_{n+1}) \times e_{a_{n+1}}
\]

\[
= (\sigma'_1 \ast \ldots \ast \sigma'_{j_{n+1}-1} \ast e_{c_1} \ast \ldots \ast e_{c_n} \ast \sigma'_{j_{n+1}+1} \ast \ldots \ast \sigma'_{n+1}) \times \tau \tau_{n+1}
\]

This establishes the claim. \( \sqcup \)

Returning To The Mainstream. Assume that \( n \) is an inner node in \( G \) with

\[
\alpha(\gamma, n) : \prod \{\gamma(k, n) \mid k \in \bullet n\} \rightarrow \tau \left( \prod \{\gamma(n, k) \mid k \in n \bullet\} \right)
\]

as its label, and assume that the edge \( \langle k, n \rangle \) is replaced by the edges \( \langle k, q \rangle, \langle q, n \rangle \) for some \( q \in Q \). The new edges are labelled through the object \( \gamma(n, q) \), and the new node \( n \) carries the label \( e_{\gamma(n, q)} \). Other edges leading into node \( n \) are also replaced. The net effect of inserting a node just in front of node \( n \) is replacing \( \alpha(\gamma, n) \) by

\[
\alpha(\gamma, n) \ast e^{\prod \{\gamma(k, n) \mid k \in \bullet n\}}
\]

which equals of course \( \alpha(\gamma, n) \). Similarly, replacing an edge \( \langle n, k \rangle \) by edges \( \langle n, q \rangle, \langle q, k \rangle \) and introducing labels on edges and on \( q \in Q \) accordingly has the effect of replacing \( \alpha(\gamma, n) \) by

\[
e^{\prod \{\gamma(n, k) \mid k \in \bullet n\}} \ast \alpha(\gamma, n),
\]
equalling \( \alpha(\gamma, n) \), too. This is a translation of the idea of inserting “neutral” nodes into the graph in order to render it stratified. In fact, two \( Q \)-extensions to \( G \) differ only by such neutral nodes on paths between nodes taken from \( G \).

46
Proposition 2.3.23 Suppose \( \langle G, \gamma \rangle \) is a PF-system with \( \langle G_1, \gamma \rangle \) and \( \langle G_2, \gamma \rangle \) as Q-extensions. Then \( P(\gamma, G_1) = P(\gamma, G_2) \).

Proof 1. The proof proceeds by induction on

\[ N := \max\{\Lambda(G_1), \Lambda(G_2)\} \]

The \( j \)th partition element of graph \( G_i \) will denoted by \( S_j^{(i)} \).

2. The induction starts at \( N = 2 \). This step inspects each node \( n \) of \( G \) in turn. Suppose \( n \in (S_1^{(1)} \setminus S_1^{(2)}) \cap V \), then, since graph \( G_2 \) is an \( Q \)-extension to \( G \), for each predecessor \( w \) of \( n \) in \( G_1 \) there exists a node \( q_w \in Q \) such that \( \langle w, q_w \rangle \) and \( \langle q_w, n \rangle \) are edges in \( G_2 \) which are labelled by the object \( \gamma(w,n) \); node \( q_w \) itself carries the label \( e_{\gamma(w,n)} \). Lemma 2.3.20 implies that the morphism \( a(\gamma, n) \) which participates in defining \( P(\gamma, G_2) \) has a factor

\[ c_{\gamma(w_1,n)} \times \cdots \times c_{\gamma(w_r,n)} \]

to the right, where \( w_1, \ldots, w_r \) are in that order all predecessors of \( n \) in \( G_1 \). A similar argument applies to \( n \in (S_1^{(2)} \setminus S_1^{(1)}) \cap V \), so that \( P(\gamma, G_1) \) differs only by factors from \( P(\gamma, G_2) \) which are identity Kleisli morphisms. Hence the assertion holds for \( N = 2 \).

3. Let \( \max\{\Lambda(G_1), \Lambda(G_2)\} = N + 1 \). We may and do assume w.l.g. that \( V \setminus (S_1^{(2)} \cup S_1^{(1)}) \neq \emptyset \), for, otherwise no node of \( V \) is directly connected to a root in either extension, so we may construct new graphs by eliminating the respective sets \( S_1 \) without changing the work of either graph.

We construct from the PF-system \( \langle G_1, \gamma \rangle \) a PF-system \( \langle G_3, \gamma \rangle \) which is an \( Q \)-extension to \( \langle G, \gamma \rangle \) such that

\[ S_1^{(3)} = \{ n \in V \mid \exists w \in W : w \rightarrow_Q^* n \text{ in } G_1 \} \cup \{ n \in V \mid \exists w \in W : w \rightarrow_Q^* n \text{ in } G_2 \}, \]

where \( \rightarrow_Q^* \) indicates that there exists a (unique) path of non-negative length that runs — with the exception of the endpoints — entirely through \( Q \). Moreover, \( P(\gamma, G_1) = P(\gamma, G_3) \) will hold.

Initially, \( \langle G_3, \gamma \rangle := \langle G_1, \gamma \rangle \). Assume \( n \in V \cap S_1^{(1)} \) such that \( n \in S_t^{(1)} \) for some \( t > 1 \). Let \( w_1, \ldots, w_r \) be all predecessors to \( n \) in \( G_1 \). Since \( G_2 \) is an \( Q \)-extension, there exist nodes \( q_1,2, \ldots, q_{1,t-1}, \ldots, q_{r,2}, \ldots, q_{r,t-1} \) in \( Q \) such that

\[
\begin{align*}
  w_1 &= q_{1,1} & \ldots & q_{1,t} &= n \\
  \vdots & \quad \vdots \\
  w_r &= q_{r,1} & \ldots & q_{r,t} &= n
\end{align*}
\]

form paths that run with the exception of their endpoints entirely through \( Q \). The edges on the \( i \)th path are labelled with the object \( \gamma(w_i,n) \) and \( a(\gamma, q_{i,j}) = c_{\gamma(w_i,n)} \).

Let \( k_1, \ldots, k_s \) be all successors to \( n \) in \( G_2 \). Remove the nodes \( \{q_{i,j} \mid 1 \leq i \leq r, 2 \leq i \leq t - 1\} \) and the edges, including \( \langle w_i, q_{i,2} \rangle \) and \( \langle q_{i,t-1}, n \rangle \) for \( 1 \leq i \leq r \) from \( G_3 \), and add nodes \( q_{1,2}', \ldots, q_{1,t-1}', \ldots, q_{s,2}', \ldots, q_{s,t-1}' \) as well as edges so that we have the paths

\[
\begin{align*}
  n &= q_{1,1}' & \ldots & q_{1,t}' &= k_1 \\
  \vdots & \quad \vdots \\
  n &= q_{s,1}' & \ldots & q_{s,t}' &= k_s.
\end{align*}
\]
Put $\gamma(q_{i,j}q_{i,j+1}) := \gamma(n,k_i)$, and set $\gamma(q_{i,j}) := \epsilon \gamma(n,k_i)$ \((i > 1, j < t)\).

Consequently, graph $G_3$ remains an $Q$-extension to $G$, and from Lemma 2.3.20, part 2, we see that $P(\gamma, G_1) = P(\gamma, G_3)$ holds. Working in this way through $V \cap (S_1^{(1)} \cup S_1^{(2)})$ will eventually produce the desired graph.

In the same manner we construct a PF-system $\langle G_4, \gamma \rangle$ with $S_1^{(4)} = S_1^{(2)}$ such that $P(\gamma, G_2) = P(\gamma, G_4)$ holds.

Remove the roots from $G_3$; this yields the graph $\tilde{G}_3$ which is an $Q$-extension to

$$\tilde{G} := (V \setminus W, E \cap (V \setminus W \times V \setminus W)).$$

Similarly, remove the roots from $G_4$ yielding $\tilde{G}_4$, which is also an $Q$-extension to $\tilde{G}$. Since $\max\{\Lambda(\tilde{G}_3), \Lambda(\tilde{G}_4)\} \leq N$, the induction hypothesis applies, so that

$$P(\gamma, G_1) = P(\gamma, G_3) = P(\gamma, \tilde{G}_3) \ast A(S_1^{(3)}, \gamma) = P(\gamma, \tilde{G}_4) \ast A(S_1^{(4)}, \gamma) = P(\gamma, G_4) = P(\gamma, G_2)$$

holds. $\dashv$

This proposition shows that the work described by an $Q$-extension of a PF-system does only depend on the underlying PF-system, so that we are now in a position to define the work of such a system — which need not be stratified — through its stratified step-twins.

**Definition 2.3.24** We define the work $P(\gamma, G)$ being done by the PF-system $\langle G, \gamma \rangle$ as the work $P(\gamma, G_1)$ of one of its $Q$-extensions $\langle G_1, \gamma \rangle$.

Consequently the work of a PF-system may be conveniently computed through one of its $Q$-extensions.

### 2.3.6 System Evolution

Our constructions support system evolution in a quite general sense. A PF-system may evolve horizontally or vertically. Horizontal evolution concatenates pipelines, with data transformations possibly serving as glue between the parts. Vertical evolution refines a pipeline by substituting a component through an entire subsystem. Both operations are vital in composing systems from smaller ones, so that larger systems can be built up through a suitable sequence of them.

**Concatenation**

Let $\langle G_1, \gamma \rangle$ and $\langle G_2, \chi \rangle$ be two PF-systems, $G_i = \langle V_i, E_i \rangle$. The idea in concatenating both is to pipe the output from the first system to the input of the second one, hence

$$V_1 \cap V_2 = B_1 = W_2$$
should hold: the output nodes from the first system should coincide with the input nodes for the second one; otherwise, these systems do not share nodes. Neither an input node nor an output node carries any functionality in our model, but by lumping them together, we may wish to perform some work (combining pipes often requires some transformation, e.g., of formats, between input and output). Hence we assume for each node $n \in B_1 = W_2$ the existence of a Kleisli morphism

$$\alpha (\tau, n) : \prod \{ \gamma (k, n) | (k, n) \in E_1 \} \to \exists \left( \prod \{ \chi (n, j) | (n, j) \in E_2 \} \right).$$

This permits defining the $\tau$-concatenation $\langle G_1, \gamma \rangle +_\tau \langle G_2, \chi \rangle$ as $\langle H, \kappa \rangle$ with

$$H := \langle V_1 \cup V_2, E_1 \cup E_2 \rangle,$$

$$\kappa (k, n) := \begin{cases} \gamma (k, n), & (k, n) \in E_1, \\ \chi (k, n), & \text{otherwise}, \end{cases}$$

$$\alpha (\kappa, n) := \begin{cases} \alpha (\gamma, n), & n \in V_1 \setminus (W_1 \cup B_1), \\ \alpha (\tau, n), & n \in B_1, \\ \alpha (\chi, n), & n \in V_2 \setminus (W_2 \cup B_2). \end{cases}$$

We get as a consequence of Proposition 2.3.23:

**Proposition 2.3.25** Under the conditions above, $\langle H, \kappa \rangle := \langle G_1, \gamma \rangle +_\tau \langle G_2, \chi \rangle$ is a PF-system, and

$$P (\kappa, H) = P (\chi, G_2) \ast \tau (\alpha (\tau, n_1) \times \ldots \times \tau \alpha (\tau, n_k)) \ast P (\gamma, G_1),$$

where $B_1 = W_2 = \{ n_1, \ldots, n_k \}$. ⊣

Thus $\tau$ provides the glue for composing the PF-systems, and the work being done exhibits the work performed when combining both systems. The glue alluded at here is different from but similar in function to the glue introduced in [96].

**Substitution**

Systems are often built through successive stages of refinements, where a part of a system is first represented as a node, and this node is then replaced in subsequent steps by an entire subsystem. This may graphically be described as glass-box refinement, see [13, 14.5].

Let $G_i = \langle V_i, E_i \rangle$ be dags with respective roots $W_i$ and leaves $B_i$, and let $n \in V_1$ be a node such that (dots taken in $G_1$)

$$\bullet n = W_2, n \bullet = B_2.$$

Thus an incoming edge for $n$ comes from a root in $G_2$, and an outgoing edge goes to a leaf in $G_2$. For technically simplifying the representation, we assume that only the nodes in $W_2 \cup B_2$ are common to $V_1$ and $V_2$. We assume further that we have a selection map

$$\psi : W_2 \cup B_2 \to V_2 \setminus (W_2 \cup B_2)$$

which will help constructing new edges when absorbing $G_2$ into $G_1$ by associating with each root or leaf an inner node as source or target of an edge, as we will see. We require
that $\psi[W_2] \cap \psi[B_2] = \emptyset$, since otherwise cycles in the replacement graph would result. Define the $\psi$-replacement

$$\mathcal{G}_1[\mathcal{G}_2 \setminus \psi n]$$

of node $n$ through graph $\mathcal{G}_2$ as the graph $(U, D)$ by

$$U := (V_1 \setminus \{n\}) \cup V_2,$$

$$D := (E_1 \cap (V_1 \setminus \{n\}))^2 \cup E_2 \cup \{\langle w, \psi(w) \rangle \mid w \in W_2\} \cup \{\langle \psi(b), b \rangle \mid b \in B_2\}.$$ 

Thus we build the new graph by combining all nodes with the exception of $n$, the node to be replaced. All edges leading into $n$ or out of it are removed, and replaced by edges into $\mathcal{G}_2$: if $\langle w, n \rangle \in E_1$ is an edge in $\mathcal{G}_1$, the node $w$ must be a root in $\mathcal{G}_2$, then this edge will be replaced in the replacement graph by the edge $\langle w, \psi(w) \rangle$, similarly for edges $\langle n, b \rangle \in E_1$. Since the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ do not have cycles, and since $\psi$ assigns by assumption different nodes to roots and to leaves, $(U, D)$ does not have any cycles either.

We apply this construction to PF-systems now. Suppose that in addition to the assumptions made so far $\langle \mathcal{G}_2, \gamma \rangle$ and $\langle \mathcal{G}_2, \chi \rangle$ are PF-systems. For getting our machinery going, the new edges need labels from $X$; these edges should not violate the typing constraints imposed on node $n$. Call the selection map $\psi$ viable iff

$$\forall w \in W_2 : \gamma(\langle w, n \rangle) = \chi(\langle w, \psi(w) \rangle) \land \forall b \in B_2 : \gamma(\langle n, b \rangle) = \chi(\langle \psi(b), b \rangle)$$

holds. This entails that the Kleisli morphisms $a(\gamma, n)$ and $P(\chi, \mathcal{G}_2)$ have the same signatures.

We define the PF-system

$$\langle \mathcal{G}_1[\mathcal{G}_2 \setminus \psi n], \gamma[\chi \setminus \psi n] \rangle$$

in the obvious way by taking the values $\gamma$, and $\chi$ for edges in $E_1$ or in $E_2$, resp., depending on where they come from, and by setting for the new edges

$$\gamma[\chi \setminus \psi n]_{\langle w, \psi(w) \rangle} := \gamma(\langle w, n \rangle),$$

similarly for $\gamma[\chi \setminus \psi n]_{\langle \psi(b), b \rangle}$. The labels for the nodes are left unchanged, coming either from $\gamma$ or from $\chi$. Then Proposition 2.3.23 implies that we may compute the work for the composed system in these steps:

- compute $P(\chi, \mathcal{G}_2)$, hence the work of the system which is to refine node $n$,
- substitute for $a(\gamma, n)$ the morphism $P(\chi, \mathcal{G}_2)$, leaving the rest of $\gamma$ alone; technically: form $\gamma[P(\chi, \mathcal{G}_2) \setminus n]$,
- compute the work done by the PF-system based on graph $\mathcal{G}_1$ with the modified value for $\gamma$.

Formally:

**Proposition 2.3.26** $P(\gamma[\chi \setminus \psi n], \mathcal{G}_1[\mathcal{G}_2 \setminus \psi n]) = P(\gamma[P(\chi, \mathcal{G}_2) \setminus n], \mathcal{G}_1)$, provided the selection map $\psi$ is viable.
Reading this equation from left to right, we see what happens when a node is substituted by an entire subsystem. Reading it from right to left it permits us to state the effect of shrinking a subsystem into a single node — this may be helpful when system evolution goes both ways, expanding nodes to subsystems, and replacing a subsystem by another one.

To illustrate: The PF-system in Figure 2.1 shall be refined as an extension to and a continuation of the discussion in section 2.3.1. Node 2 will be replaced by the subsystem depicted in Figure 2.5.

Assign relations with the following signatures to the innernodes:

\[ R_A \subseteq (X_2 \times X_3) \times (Y_1 \times Y_2), \]
\[ R_B \subseteq Y_1 \times X_5, \]
\[ R_C \subseteq Y_2 \times X_6. \]

The work being done by this subsystem is then given by

\[ \{ \langle x_2, x_3, x_5, x_6 \rangle \mid \exists y_1 \in Y_1, y_2 \in Y_2 : \langle x_2, x_3, y_1, y_2 \rangle \in R_A, \langle y_1, x_5 \rangle \in R_B, \langle y_2, x_6 \rangle \in R_C \}. \]

The reader is invited to formulate \( \langle G_2, \chi \rangle \) and a viable map \( \psi \). Proposition 2.3.26 then gives the work of the entire system, when the replacement has been done, as

\[ \{ \langle x_1, x_2, x_7, x_8 \rangle \mid \exists x_3 \in X_3, x_4 \in X_4, x_5 \in X_5, x_6 \in X_6 \exists y_1 \in Y_1, y_2 \in Y_2 : \]
\[ \langle x_1, x_3, x_4 \rangle \in R_1, \langle x_2, x_3, y_1, y_2 \rangle \in R_A, \langle y_1, x_5 \rangle \in R_B, \langle y_2, x_6 \rangle \in R_C \]
\[ \langle x_4, x_5, x_7 \rangle \in R_3, \langle x_6, x_8 \rangle \in R_4 \}. \]

The resulting system is given in Figure 2.6.

For the stochastic case one assumes similarly that stochastic relations

\[ K_A : \quad X_2 \times X_3 \leadsto H \times Y_1 \times Y_2, \]
\[ K_B : \quad Y_1 \leadsto H \times X_5, \]
\[ K_C : \quad Y_2 \leadsto H \times X_6 \]

are assigned to the subsystem’s nodes. In view of Proposition 2.3.19 the monoid \( H \) should be commutative. The work is computed through the Kleisli product, yielding an expression for the measure \( K_2(x_2, x_3) \) in terms of \( K_A, K_B \) and \( K_C \) which is then substituted.
into the integral elaborated in Example 2.3.12. Specifically, the subsystem’s work is described by the stochastic relation \( K_s : X_2 \times X_3 \rightsquigarrow H \times X_5 \times X_6 \) such that for \( Q \subseteq H \times X_5 \times X_6 \)

\[
K_s(x_2, x_3)(Q) = \int_{H \times Y_1 \times Y_2} (K_B(y_1) \otimes_H K_C(y_2)) \left( \{ \langle h, x_5, x_6 \rangle \mid \langle h_a h, x_5, x_6 \rangle \in Q \} \times K_A(x_1, x_2)(d\langle h_a, y_1, y_2 \rangle) \right) \\
= \int_{H \times Y_1 \times Y_2} \int_{H \times X_6} K_B(y_1) \left( \{ \langle h_b, x_5 \rangle \mid h_a h_c h_b, x_5, x_6 \in Q \} \times K_C(y_2)(d\langle h_c, x_6 \rangle) K_A(x_1, x_2)(d\langle h_a, y_1, y_2 \rangle). \right)
\]

Consequently, we have for \( f : H \times X_5 \times X_6 \to \mathbb{R} \) measurable and bounded

\[
\int_{H \times X_5 \times X_6} f(h, x_5, x_6) \, dK_s(x_2, x_3)(d\langle h, x_5, x_6 \rangle) \\
= \int_{H \times Y_1 \times Y_2} \int_{H \times X_6} \int_{H \times X_5} f(h_a h_b h_c, x_5, x_6) K_B(y_1)(d\langle h_b, x_5 \rangle) \times K_C(y_2)(d\langle h_c, x_6 \rangle) K_A(x_1, x_2)(d\langle h_a, y_1, y_2 \rangle). 
\]

Note the accumulating behavior of the monoid’s elements.

This integral is needed, because the final equation describing the system’s work before replacement in Example 2.3.12 requires us to describe integration with respect to \( K_2(x_1, x_3) \). Substituting, we get from that equation

\[
\int_{H \times Y_3 \times Y_4} \int_{H \times Y_1 \times Y_2} \int_{H \times X_4} \int_{H \times X_5} \int_{H \times X_7} K_4(x_6)\{\langle h_1, x_8 \rangle \mid \langle g_1(h_a h_c h_b) g h_1, x_7, x_8 \rangle \in F\} K_3(x_4, x_5)(d\langle g, x_7 \rangle) \times K_B(y_1)(d\langle h_b, x_5 \rangle) K_C(y_2)(d\langle h_c, x_6 \rangle) K_A(x_1, x_2)(d\langle h_a, y_1, y_2 \rangle) \times K_1(x_1)(d\langle g_1, x_3, x_4 \rangle).
\]
This formidable expression describes the probability that input \((x_1, x_2)\) produces an element of the Borel set \(F \subseteq H \times X_7 \times X_8\). It is interesting to see how each component leaves its trace in the result’s monoid compartment.

### 2.3.7 Related Approaches

Modelling of pipelines through the specification language Z summarized and discussed e.g. in \([85, 82, 1]\) is evidently much closer to an implementation than the approach proposed in the present paper. Thus a person intending to implement a system with such an architecture is probably better off looking at a Z-specification, using well-known refinement techniques like the ones discussed by Spivey \([87, \text{Ch. 5}]\) for coming even closer to a realization as a running system.

The difference to the present approach, however, lies deeper: Shaw at al. emphasize the first class rank of architectural connectors \([85, 1, 82, 83, 84]\). This implies that filters and pipes are treated on the same eye-level. The scenario here marks a contrast: connectors are represented through objects in a category, components through morphisms of a rather special kind, putting these two kinds of entities on different levels. It may much be said in favor of dealing uniformly with connectors and with components, but it seems that an asymmetric treatment helps the intuition: computations are conceptually different from data transport, however complex the latter may be. The present approach reflects an approach like “Tell me, what your data are, then we will talk about computations on them”, very similar to the one in object-oriented software construction.

The approach proposed by Fiadeiro et al, see e.g. \([96, 34]\), using categories for modelling architectures shows how different kinds of functors, in particular interface functors, may be put to use for constructing systems. This is illustrated in \([96]\) where a diagram is “compiled” through computing its colimit, leading to the early version of a program. Moreover, fundamental kinds of interactions of program components are studied using the patterns constructed in that paper. The focus lies on modelling just the interactions for a particular class of mobile programs, emphasizing the importance of connectors: “Software Architecture has put forward the concept of connector to express complex relationships between system components, thus facilitating the separation of coordination from computation.” is the very first sentence in the paper’s abstract. The computation proper, however, has not been addressed, and this is what we propose in the present paper. Reflecting the mobile nature of the programs discussed in \([96]\), and taking into account that no fixed topology is available for the computing nodes in such a scenario, another difference becomes visible: the topology of the communication and the direction of data flow remains fixed here but may be subject to change dynamically in a mobile context. But this is a completely different story, since PF-systems exhibit a fixed structure by their very nature.

Program evolution is supported by concatenating, and by hierarchically composing PF-systems. While the first operation is easily modelled in the Z-approach, only a hint at supporting the hierarchical composition is given in \([1]\), making it difficult to compare both approaches in this respect. The Reo calculus introduced by Arbab and Rutten is based on timed data streams; it supports a topology of connections that is inherently dynamic, and it is based on a distinctive separation of data and time. Examples exhibit the intriguing yet powerful simplicity of this calculus (which also may accommodate mobility). Because we focus in the present
work on a fixed communication topology, a comparison between Reo and the constructions proposed here shows that Reo is both more general and more specific: Reo focusses on set-theoretic relations only, monads are not mentioned. Timing plays a crucial role, so timed sequences are used. On the other hand, arbitrary relations are allowed, so are arbitrary compositions, which permits coinduction as the guiding principle for proofs. Finally, the calculus has all kinds of channels, not only pipes. The central difference is, however, that coordination forms a first class concept in this approach.

Barbosa [8] considers components, i.e., state-based dynamical systems. He models them as coalgebras for a class of endofunctors in the category of sets and shows how to obtain a behavioral model through parametrizations by a strong monad. This permits him to capture some important behavioral features like partiality or nondeterminism. The guiding proof principle is a variant of coinduction. Probabilistic settings are outside this rich thesis.

The FOCUS approach due to M. Broy and K. Stølen outlined in [13] provides specification, refinement and verification techniques for the development of interactive systems, thus is not tied to one particular architecture. The basic construction is that of a timed stream of data, refinement is the basic architectural operation, where different kinds of refinements are investigated. For example, Broy and Stølen discuss glass-box refinements which is the exact counterpart to the substitution operation investigated in section 2.3.6, and they discuss composing systems through an operator style [13, 14.5 resp. 5.3.3]. FOCUS has a much broader spectrum of application than the present approach. This is due to the fact that no particular architecture is aimed at; there is, however, no explicit notion of computation there, while the present approach encapsulates computations through monads.

Possible Extensions

The present discussion excludes those PF-systems that are cyclic by discussing pipelines only; further work should admit systems that contain cycles, and, more generally, by admitting decisions which component to invoke next. Here the extensive categories investigated e.g. in [16, 91] are be helpful. Extensive categories stress the use of pullbacks somewhat, on the other hand, Corollary 4.3.7 will show that pullbacks in the category of stochastic relations with surjective measurable maps as morphisms do not exist. A first and very encouraging step towards investigating layered architectures within the framework of extensive categories can be found in [57].

Introducing timing and explicit synchronization is another area that needs further consideration — the model proposed here is abstract enough not to unduly constrain the modeler, but on the other hand some support could be offered, even at the price of restricting the model. As far as properties of the models are concerned, proof rules which permit stating properties for systems that are evolving according to section 2.3.6 are of interest. For a different probabilistic approach to synchronization using automata see [19, 15].

It is challenging to see how other architectural styles are tackled, and how to model dynamically changing communication topologies. On the relational side, we have narrowed down monads which represent the two major kinds of relations. The natural transformations \( \theta \) and the \( \sharp \)-condition seem to be ingredients to a monad which models relations.
2.4 Case Study: Probabilistic Semantics of a Simple Language

While section 2.3 discusses a typical problem arising from programming in the large, emphasizing a uniform approach for nondeterministic and stochastic relations by formulating the model in a common framework, we will deal now with a problem of programming in the small and investigate the semantics of a simple programming language. We focus in this discussion not on a common approach (nondeterministic approaches to programming language semantics are very well known, and have been formalized through e.g. modal logics), but show rather that a program can be perceived as a device that transforms probabilities. This approach is known as well for quite some time [55] and has even been used for the average case analysis of algorithms [24]. We extend these ideas by formulating the state space of a program as the stochastic relations over some space. As a case in point, we will first give a simple language, a partial correctness semantics and a partial correctness logic for it. Both, semantics and logic, will work on the basis of transforming continuous stochastic relations. We demonstrate consistency and completeness, so that we are on safe grounds when dealing with the language’s semantics. When dealing with the semantics of the language, we will have to deal with the semantics of the while-loop. This is captured through a fixed point, traditionally through an application of the Kleene-Knaster-Tarki Theorem or one its many refinements. But there is no complete partial order visible at this point, so we have to find other means of securing the existence of a fixed point. Since we are working in the realm of a complete metric space, on which continuous functions operate, it may be tempting to apply Banach’s famous Theorem, but, alas, there is no contraction visible on the horizon (see [63] for an application of the Banach Theorem to semantics). Thus we have to follow a custom tailored path and find a fixed point on our own. It will turn out that projective limits are the tool to use.

We will first define the language Ludwig, prepare for the semantics and formulate the fixed point argument, and then give a partial correctness semantics and a partial correctness logic. Consistency and completeness is established as an important property. This topic is investigated further when discussing derandomization in section 3.7.

2.4.1 The Language: Ludwig

The language Ludwig\(^2\) has assignments, conditional statements, loops and composition. Its syntax is given through:

\[
\beta ::= \text{skip} \mid a := E \mid \text{if } \alpha \text{ then } \beta \text{ else } \beta \text{ fi } \mid \text{while } \alpha \text{ do } \beta \text{ od } \mid \beta;\beta
\]

(we will abbreviate the conditional command by \(\text{if } \alpha \text{ then } \beta \text{ fi}\) if the alternative branch is \(\text{skip}\)). The variables \(a\) are taken from a set \(\text{Vars}\), the expressions \(E\) from a set \(\text{Exprs}\). Let \(X\) be a Polish space which will be fixed in the sequel. A Boolean \(\alpha\) is modelled through a map \(\alpha : X \rightarrow \{\text{tt}, \text{ff}\}\) from \(X\) to the truth values \(\{\text{tt, ff}\}\). It is natural to assume

\(^2\)Ludwig Wittgenstein writes in his *Philosophische Untersuchungen*: “Wer in ein fremdes Land kommt, wird manchmal die Sprache der Einheimischen durch hinweisende Erklärungen lernen, die sie ihm geben; und er wird die Deutung dieser Erklärungen oft raten müssen und manchmal richtig, manchmal falsch raten” [98, Nr. 32, p.29]
measurability of this map, so that the set
\[
\{ \alpha = \text{tt} \} := \{ x \in X \mid \alpha(x) = \text{tt} \}
\]
is a Borel set (and \( \{ \alpha = \text{ff} \} \) as well). Because there are at most countably many of these sets for Ludwig, we can find a Polish topology on \( X \) which is finer than the given one but having the same Borel sets such that these sets are clopen, see Proposition A.2.2. We will assume that \( X \) is equipped with this topology.

We assume that we have a map
\[
\mathcal{E} : \text{Vars} \to (\text{Exprs} \to (X \to \mathcal{S}(X)))
\]
at our disposal so that for each variable \( a \) and each expression \( E \) the continuous stochastic relation \( \mathcal{E}(a)(E) \) is continuous. The map \( \mathcal{E} \) is intended to model the assignment of \( E \) to variable \( a \) by evaluating the environment, performing the computation which is associated with the expression \( E \) and assigning the result to \( a \).

We need some preparations for defining the semantics of Ludwig.

### 2.4.2 Some Preparations and a Fixed Point

Associate with the Boolean \( \alpha \) the continuous stochastic relations \( \chi^{\alpha}_{\{\alpha=\text{tt}\}} \) and \( \chi^{\alpha}_{\{\alpha=\text{ff}\}} \). These relations are defined through
\[
\chi^{\alpha}_Z(x)(A) := \chi_Z(x) \cdot \delta_x(A),
\]
where as usual \( \chi_Z \) is the indicator function of set \( Z \), and \( \delta_x \) is the Dirac measure on \( x \). If \( Z \subseteq X \) be a clopen set, then \( \chi^{\alpha}_Z \) is a continuous stochastic relation. This is so since for continuous \( f : X \to \mathbb{R} \) the map
\[
x \mapsto \int_X f \, d\chi^{\alpha}_Z(x) = f(x) \cdot \chi(x)
\]
is continuous. It will be convenient to have some easy rules for manipulating \( \chi^{\alpha}_Z \) at our disposal.

**Lemma 2.4.1** Let \( \ell : X \to X \) be a continuous stochastic relation, \( Z \subseteq X \) be a clopen set, and \( f : X \to \mathbb{R} \) be a measurable and bounded map. Then

1. \( \int_X f \, d\left( \chi^{\alpha}_Z * \ell \right)(x) = \int_Z f \, d\ell(x) \), so that in particular \( (\chi^{\alpha}_Z * \ell)(x)(A) = \ell(x)(Z \cap A) \) holds for a Borel set \( A \subseteq X \).
2. \( \int_X f \, d\left( \ell * \chi^{\alpha}_Z \right)(x) = \chi_Z(x) \cdot \int_X f \, d\ell(x) \), so that in particular for a Borel set \( A \subseteq X \)
   \( (\ell * \chi^{\alpha}_Z)(x)(A) = \chi_Z(x) \cdot \ell(x)(A) \) holds.
3. Both \( (\ell * \chi^{\alpha}_Z) \) and \( (\chi^{\alpha}_Z * \ell) \) are continuous stochastic relations.

**Proof** 1. If \( A \in \mathcal{B}(X) \), then
\[
(\chi^{\alpha}_Z * \ell)(x)(A) = \int_X \chi_Z(y) \cdot \delta_y(A) \, \ell(x)(dy) = \ell(x)(Z \cap A),
\]
hence the equality holds for \( f = \chi_A \) with \( A \in \mathcal{B}(X) \). Linearity of the integral warrants that the equality holds for non-negative step functions, i.e. for functions \( f \) of the form \( f = \sum_{i=1}^{n} a_i \cdot \chi_{A_i} \) with \( A_i \in \mathcal{B}(X) \) and non-negative coefficients \( a_i \).

Since for each measurable and bounded \( f \geq 0 \) there exists a monotone increasing sequence of step functions that has \( f \) as its supremum, and since the integral is continuous with respect to bounded monotone convergence, the equality holds for bounded and measurable \( f \geq 0 \). Because each bounded and measurable function can be decomposed as the difference of two bounded, measurable and non-negative functions, and since the integral is additive, the claim is established for the first part. The second part is proved exactly in the same way.

2. The third part is an easy consequence of the first two parts, because \( Z \) is a clopen set, so that \( \chi_Z \) is a continuous function. \( \dashv \)

Assume \( \ell : X \rightsquigarrow X \) is a continuous stochastic relation, and define inductively

\[
\ell^{(1)}(x) := \ell(x) \\
\ell^{(k+1)}(x)(A) := \int_X \ell^{(k)}(x_1)\{\{x_2, \ldots, x_{k+1}\} \mid \{x_1, \ldots, x_{k+1}\} \in A\}\ell(x)(dx_1), \\
(A \in \mathcal{B}(X^{k+1})).
\]

This construction is not unlike the one that leads to the projective limit, as a comparison with the construction at the end of section A.3.3 shows. In fact, if always \( \ell(x)(X) = 1 \) holds, then we may obtain the projective limit from it. This will be helpful in the proof of Lemma 2.4.3, where we construct a projective limit that serves as an upper bound.

Motivating the definition of \( \ell^{(k)} \), suppose that \( \ell \) governs the move of a particle in one step, then \( \ell^{(k)}(x) \) describes what happens when the particle starts at \( x \) and does \( k \) steps. It is this construction that will permit us to trace the dynamics of a while-loop in Ludwig. The next Lemma fixes some technical properties. We use the notation \( \ell^k \) for the \( k \)th iterate of \( \ell \) (so \( \ell^1 = \ell \), \( \ell^{k+1} = \ell^k \circ \ell \); notice that \( \ell^k \neq \ell^{(k)} \)).

**Lemma 2.4.2** Let \( \ell : X \rightsquigarrow X \) be a continuous stochastic relation, then

1. \( \ell^{(k)} : X \rightsquigarrow X^k \) is a continuous stochastic relation for each \( k \in \mathbb{N} \).

2. \( \ell^{(k+1)}(x)(A \times X) \leq \ell^{(k)}(x)(A) \) holds for each Borel set \( A \subseteq X^k \). If \( \ell(x)(X) = 1 \) for all \( x \in X \), then \( \ell^{(k+1)}(x)(A \times X) = \ell^{(k)}(x)(A) \) for all Borel sets \( A \in \mathcal{B}(X^k) \).

3. For each \( k \in \mathbb{N} \) and each Borel set \( A \in \mathcal{B}(X) \) we have

\[
\left( \ell \ast \chi_Z^k \right)^k(x)(A) = \chi_Z(x) \cdot \ell^{(k)}(x)(Z^{k-1} \times A).
\]

**Proof** 1. It is clear that \( \ell^{(k)}(x) \) defines for each \( k \in \mathbb{N} \) a sub-probability measure, so the crucial point will be to establish that \( x \mapsto \ell^{(k)}(x) \) is weakly continuous. This is established by induction on \( k \), the induction begin at \( k = 1 \) being trivial. The induction step takes a continuous and bounded function \( f : X^{k+1} \to \mathbb{R} \), then the same method as applied in the proof for part 1 of Lemma 2.4.1 shows that

\[
\int_{X^{k+1}} f \; d\ell^{(k+1)}(x) = \int_X \left( \int_{X^k} f(x_1, \ldots, x_{k+1}) \; d\ell^{(k)}(d[x_2, \ldots, x_{k+1}]) \right) \ell(x)(dx_1)
\]
(as in the proof of Lemma 2.4.2 one establishes the equation first for \( f = \chi_A \) with \( A \in \mathcal{B}(X^{k+1}) \), then approximates through step functions for \( f \) positive, and decomposes general \( f \) into a positive and a negative part; it is not important for this argument that \( f \) is continuous, measurability suffices).

The induction hypothesis and the continuity of \( f \) yields that

\[
x_1 \mapsto \int_{X^k} f(x_1, \ldots, x_{k+1}) \, d\ell^{(k)}(d\langle x_2, \ldots, x_{k+1} \rangle)
\]

constitutes a continuous function, which implies by the continuity of \( \ell \) that

\[
x \mapsto \int_{X^{k+1}} f \, d\ell^{k+1}(x)
\]

is continuous as well.

2. The second assertion is also established by induction on \( k \), the crucial step being the begin at \( k = 2 \). We assume first that \( \ell(x)(X) = 1 \) for all \( x \in X \). Let \( A \in \mathcal{B}(X) \), then

\[
\ell^{(2)}(x)(A \times X) = \int_X \ell(x_1)(\{x_2 \mid (x_1, x_2) \in A \times X\}) \, \ell(x)(dx_1)
\]

\[
= \ell(x)(A).
\]

The equality for the induction step follows immediately from the induction hypothesis. It can be read off the definition of \( \ell^{(k+1)} \), using the assumption that \( \ell(x)(X) = 1 \) always holds.

The general case \( \ell(x)(X) \leq 1 \) is proved in exactly the same way: the induction step then reads

\[
\ell^{(k+1)}(A \times X) = \int_X \ell^{(k)}(x_1)(\{x_2, \ldots, x_{k+1} \mid (x_1, \ldots, x_{k+1}) \in A \times X\}) \, \ell(x)(dx_1)
\]

\[
\leq \int_X \ell^{(k-1)}(x_1)(\{x_2, \ldots, x_k \mid (x_1, \ldots, x_k) \in A\}) \, \ell(x)(dx_1)
\]

\[
= \ell^{(k)}(x)(A).
\]

3. We prove the third assertion by induction as well. The begin at \( k = 1 \) is trivial, the induction step makes use of Lemma 2.4.1:

\[
(\ell \ast \chi_{Z}^{k+1})(x)(A) = \int_{Z} (\ell \ast \chi_{Z}^{k})(x_1)(A) \, (\ell \ast \chi_{Z}^{1})(x)(dx_1)
\]

\[
= \chi_{Z}(x) \cdot \int_{Z} \ell^{(k)}(x_1)(Z^{k-1} \times A) \, \ell(x)(dx_1)
\]

\[
= \chi_{Z}(x) \cdot \int_{X} \ell^{(k)}(x_1)(\{a \in X^{k} \mid (x_1, a) \in Z \times Z^{k-1} \times A\}) \, \ell(x)(dx_1),
\]

and from this follows the assertion. \( \dashv \)

**Lemma 2.4.3** Let \( \ell : X \rightharpoonup X \) be a continuous stochastic relation, and assume that \( Z \subseteq X \) is a clopen set. Put

\[
\ell_{Z,\infty}(x)(A) := \sum_{k=1}^{\infty} (\ell \ast \chi_{Z}^{k})(x)(A \cap X \setminus Z)
\]

for the Borel set \( A \subseteq X \). Then \( \ell_{Z,\infty} : X \rightharpoonup X \) defines a continuous stochastic relation.
2.4 Case Study: Probabilistic Semantics of a Simple Language

Proof 1. We can always find a continuous stochastic relation $\tilde{\ell}$ such that for all $x \in X$ both $\ell(x) \leq \tilde{\ell}(x)$ and $\tilde{\ell}(x)(X) = 1$ holds. We make use of the sequence $\tilde{\ell}^{(k)}$ of continuous stochastic relations defined above: Corollary A.3.9 that there exists exactly one probability measure $\tilde{\ell}^{(\infty)}(x)$ on the Borel sets of the infinite product $X^\infty := \prod_{k \in \mathbb{N}} X$ such that

$$\tilde{\ell}^{(\infty)}(x)(A \times \prod_{j > k} X) = \tilde{\ell}^{(k)}(x)(A)$$

holds for each Borel sets $A \in \mathcal{B}(X^n)$. This is so since the sequence $\left(\tilde{\ell}^{(k)}(x)\right)_{k \in \mathbb{N}}$ forms a projective system by part 2 in Lemma 2.4.2: $\tilde{\ell}^{(k)}(x)$ is the image of $\tilde{\ell}^{(k+1)}(x)$ under the projection to the first $k$ components.

2. Expanding the definition of $\ell_{Z,\infty}$ and using the representation of part 3 in Lemma 2.4.2 we see that

$$\ell_{Z,\infty}(x)(A) = \sum_{k=1}^{\infty} \left(\ell \ast \chi^Z_k\right)^k(x)(A \cap X \setminus Z) \geq \sum_{k=1}^{\infty} \chi_Z(x) \cdot \tilde{\ell}^{(k)}(x)(Z^{k-1} \times (A \cap X \setminus Z)) \geq \chi_Z(x) \cdot \sum_{k=1}^{\infty} \tilde{\ell}^{(\infty)}(x)(Z^{k-1} \times (A \cap X \setminus Z) \times \prod_{j > k} X) \geq \chi_Z(x) \cdot \tilde{\ell}^{(\infty)}(x)(X^\infty) \leq 1.$$  

3. Continuity of $\ell_{Z,\infty}$ is proved through the Portmanteau Theorem A.3.2 by showing that

$$\liminf_{n \to \infty} \ell_{Z,\infty}(x_n)(G) \geq \ell_{Z,\infty}(x_0)(G)$$

whenever $x_n \to x_0$ and $G \subseteq X$ is an open set. In fact, by Fatou’s Lemma [10, Theorem 16.3] and the continuity of all $\ell^{(k)}$ we see

$$\liminf_{n \to \infty} \ell_{Z,\infty}(x_n)(G) \geq \sum_{k \geq 0} \liminf_{n \to \infty} \left(\ell \ast \chi^Z_k\right)^{(k)}(x_n)(G) \geq \sum_{k \geq 0} \left(\ell \ast \chi^Z_k\right)^{(k)}(x_0)(G) = \ell_{Z,\infty}(x_0)(G),$$

establishing the desired inequality. \[ \]

The construction of the projective limit $\tilde{\ell}^{(\infty)}$ in Lemma 2.4.3 has been necessary in order to provide a measure so that the mutually disjoint summands can be subsumed under one single sum. Only then it can be shown that this sum does not exceed one. A slightly more compact representation of $\ell_{Z,\infty}$ is

$$\ell_{Z,\infty} = \sum_{k=1}^{\infty} \chi^Z_{X \setminus Z} \ast (\ell \ast \chi^Z_k)^k;$$

59
when applying part 1 of Lemma 2.4.1.
Having a look at $\ell_{Z,\infty}(x)(A)$, we see that this is the probability for an element of $Z$ to move through elements of $Z$ until at some finite point in time an element of $A$ which must not be in $Z$ is encountered. Put $Z = \{ \alpha = \text{tt} \}$, then it becomes clear that $\ell_{\{\alpha = \text{tt}\},\infty}$ will a prime candidate from which the semantics of the while-loop is modelled. This is also due to the fact that $\ell_{\{\alpha = \text{tt}\},\infty}$ in the special case mentioned above is a projective limit, which permits to describe finite behavior through finite approximations like $\ell^{(k)}$ without having to reach for the limit at infinity.

We will associate to each program statement a continuous stochastic relation; the states of the program will be modelled through continuous stochastic relations as well. We have seen that these relations are a Kleisli construction for the Giry monad. As pointed out in [66, 48, 49], the monad may be obtained in an alternative way (what is called in [49, §4] the Kleisli point of view) through an extension of a Kleisli morphism $K : X \rightarrow \mathcal{G}(Y)$ to a morphism $K^\bullet : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$.

**Lemma 2.4.4** Let $K : X \sim Y$ be a continuous stochastic relation for the Polish spaces $X$ and $Y$, and define for $\mu \in \mathcal{G}(X), A \in \mathcal{B}(Y)$

$$K^\bullet(\mu)(A) := \int_X K(x)(A) \mu(dx).$$

Then $K^\bullet : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$ is continuous.

**Proof** Let $(\mu_n)_{n \geq 0}$ be a sequence in $\mathcal{G}(X)$ with $\mu_n \rightharpoonup w \mu_0$, and let $f : Y \rightarrow \mathbb{R}$ be continuous and bounded. We need to show that $\int_Y f dK^\bullet(\mu_n)$ converges to $\int_Y f dK^\bullet(\mu_0)$, as $n \rightarrow \infty$. We infer from the definition that

$$\int_Y f dK^\bullet(\mu_n) = \int_X \left( \int_Y f dK(x) \right) \mu_n(dx)$$

holds (repeating e.g. the approximation procedure as in the proof of Lemma 2.4.1 after having inferred that the equality above for $f = \chi_A$ is but a reformulation of the definition). Since $K : X \rightarrow \mathcal{G}(Y)$ is continuous, we see that

$$x \mapsto \int_Y f dK(x)$$

is continuous and bounded, so that

$$\int_Y f dK^\bullet(\mu_n) \rightarrow \int_Y f dK^\bullet(\mu_0)$$

follows, establishing the claim. $\dashv$ Consequently, $K^\bullet$ can be extended so that it maps continuous stochastic relations to continuous stochastic relations, upon setting $K^\bullet(L)(z)(A) := K^\bullet(L(z))(A)$, whenever $L : Z \sim X$ for an arbitrary Polish space $Z$, reusing the symbol $K^\bullet$. Consequently,

$$K^\bullet(L)(z)(A) = K^\bullet(L(x))(A) = \int_X K(x)(A) L(z)(dx) = (K \ast L)(z)(A).$$
Thus $K^\star(L) = K \ast L$, an equality which is well known, and which will be helpful in a moment.

The semantics for the while-loop will be modelled through a fixed point, using this projective limit. Computing a fixed point, one would traditionally resort either to the Kleene-Knaster-Tarski construction or to a variant of the Banach fixed point theorem [63]. The former version requires a continuous map in a complete lattice, the latter one works for a contraction in a complete metric space. Neither approach is feasible for the situation at hand: an appropriate order relation is not available, and a contracting map does not suggest itself. Thus the fixed point is determined directly through a kind of limiting process.

**Proposition 2.4.5** Put for $\ell : X \rightsquigarrow X$ and the clopen set $Z \subseteq X$

$$\rho_Z(\ell) := \chi_{X \setminus Z}^\sharp + \ell_{Z,\infty}.$$  

Then $\rho_Z(\ell)$ is idempotent with respect to composition.

**Proof** We need to establish that

$$\rho_Z(\ell) \ast \rho_Z(\ell) = \rho_Z(\ell)$$

holds. Because the composition of stochastic relations is left and right distributive, we obtain

$$\rho_Z(\ell) \ast \rho_Z(\ell) = \left(\chi_{X \setminus Z}^\sharp + \ell_{Z,\infty}\right) \ast \left(\chi_{X \setminus Z}^\sharp + \ell_{Z,\infty}\right) = \chi_{X \setminus Z}^\sharp \ast \chi_{X \setminus Z}^\sharp + \chi_{X \setminus Z}^\sharp \ast \ell_{Z,\infty} = \chi_{X \setminus Z}^\sharp + \ell_{Z,\infty} = \rho_Z(\ell).$$

This is so since $\chi_{X \setminus Z}^\sharp \ast \chi_{X \setminus Z}^\sharp = \chi_{X \setminus Z}^\sharp$ and $\chi_{X \setminus Z}^\sharp \ast \chi_{Z}^\sharp = 0$ so that both $\ell_{Z,\infty} \ast \chi_{X \setminus Z}^\sharp$ and $\ell_{Z,\infty} \ast \ell_{Z,\infty}$ vanish.

Note that $\ell_{Z,\infty}$ is an infinite sum. The above identities hold for the finite sums, and a passage to the limit maintains them, since composition is a continuous operation; this justifies the shorthand treatment, so that there is no need to go into the somewhat lengthy but not particularly entertaining details. ⊣

Interpreting $\rho_Z(\ell)$ as a map, we find that it yields a fixed point.

**Corollary 2.4.6** Given $\ell : X \rightsquigarrow X$ and the clopen set $Z \subseteq X$, $\rho_Z(\ell)^\star(g)$ is a fixed point for $\rho_Z(\ell)^\star$, where $g : X \rightsquigarrow X$ is any continuous stochastic relation.

**Proof** We obtain from Proposition 2.4.5

$$\rho_Z(\ell)^\star(\rho_Z(\ell)^\star(g)) = \rho_Z(\ell) \ast \rho_Z(\ell) \ast g = \rho_Z(\ell) \ast g = \rho_Z(\ell)^\star(g).$$

⊣

Now put

$$\rho_\alpha(\ell) := \rho_{\{\alpha=tt\}}(\ell)$$

61
for the Boolean $\alpha$ and $\ell : X \nrightarrow X$. Understanding $(\ell \ast \chi^{\sharp}_{\{\alpha=tt\}})^0$ as the identity, we may write then $\rho_\alpha(\ell)$ in a more concise form as

$$\rho_\alpha(\ell) = \sum_{k=0}^\infty \chi^{\sharp}_{\{\alpha=ff\}} \ast (\ell \ast \chi^{\sharp}_{\{\alpha=tt\}})^k,$$

indicating that we go through $k$ iterations applying $\ell$ with $\alpha = tt$ (including the case $k = 0$) before we hit $\alpha = ff$ with a noop.

This fixed point may be characterized in a different way that is more useful for the purposes below.

**Corollary 2.4.7** Let $\ell : X \nrightarrow X$ be a continuous stochastic relation, and $Z \subseteq X$ be clopen. Then $\rho_Z(\ell)$ is the only continuous solution to the equation

$$\zeta = \chi^{\sharp}_{X \setminus Z} + \zeta \ast (\ell \ast \chi^{\sharp}_{Z}),$$

provided $\inf_{k \in \mathbb{N}} \ell^{(k)}(x)(Z^k) = 0$ for all $x \in X$.

**Proof** 1. Inserting the definition of $\rho_Z(\ell)$ into the equation above shows that this is actually a solution.

2. Expanding the equation yields by an easy inductive proof

$$\zeta = \sum_{k=0}^r \chi^{\sharp}_{X \setminus Z} \ast \left( (\ell \ast \chi^{\sharp}_{Z})^k \ast \zeta \ast (\ell \ast \chi^{\sharp}_{Z})^{r+1} \right)(x)(A) = \int_Z \zeta(y)(A) (\ell \ast \chi^{\sharp}_{Z})^{k+1}(x)(dy) \leq (\ell \ast \chi^{\sharp}_{Z})^{k+1}(x)(X) = \ell^{(k+1)}(x)(Z^k \times X) = \ell^{(k)}(x)(Z^k) < \epsilon.$$

Thus the term $\zeta \ast (\ell \ast \chi^{\sharp}_{Z})^{r+1}$ vanishes, as $r \to \infty$, giving in fact $\zeta = \rho_Z(\ell)$.

The condition in the Corollary above may be reformulated, if $\ell(x)(X) = 1$ holds for all $x \in X$, because in this case we have a projective limit $\ell^{(\infty)}$: Since

$$\prod_{k \in \mathbb{N}} Z = \bigcap_{k \in \mathbb{N}} \left( Z^k \times \prod_{j > k} X \right),$$

we obtain from the construction of $\ell^{(\infty)}$ (see the part 1 of the proof for Lemma 2.4.3)

$$\ell^{(\infty)}(x)(\prod_{n \in \mathbb{N}} Z) = \inf_{k \in \mathbb{N}} \ell^{(\infty)}(x)(Z^k \times \prod_{j > k} X) = \inf_{k \in \mathbb{N}} \ell^{(k)}(x)(Z^k).$$

Thus an infinite path starting at state $x$ will stay always in $Z$ with probability 0. This will be helpful when discussing termination of Ludwig programs.
2.4.3 Partial Correctness Semantics

A state is a continuous stochastic relation, and the semantics will model state transformations. Thus we perceive a program as a device which produces a continuous stochastic relation from another one, in this way performing a transformation between states. More formally, we will show how a semantic function $C[\cdot]$ operates on these stochastic relations. We associate with each program statement $\beta$ a continuous stochastic relation, denoted by $|\beta|$, subject to the rules outlined below. If $g$ denotes the current state, then the new state after executing $\beta$ will be $|\beta| \cdot (g)$, as defined after Lemma 2.4.4. We abbreviate $|\beta| |\alpha| = \text{tt} := |\beta| \cdot \chi_{\{\alpha = \text{tt}\}}$, similarly $|\beta| |\alpha| = \text{ff} := |\beta| \cdot \chi_{\{\alpha = \text{ff}\}}$.

Termination needs to be discussed. The while-loop

$\text{while } \alpha \text{ do } \beta \text{ od}$

terminates iff $\beta$ terminates, and if there do not exist infinite paths $\langle x_0, x_1, \ldots \rangle$ such that $\alpha(x_k) = \text{tt}$ always holds. This implies that

$$\inf_{k \in \mathbb{N}} |\beta|^k(x)(\{\alpha = \text{tt}\})^k = 0$$

holds for each $x \in X$ in terms of the continuous stochastic relation $|\beta|$ associated with program statement $\beta$. For the other statements termination is defined recursively. The rules below give the semantic function.

1. Clearly, skip works as the monad's identity:

$$C[\text{skip}] := \epsilon_X,$$

thus $(C[\text{skip}]g)(x)(A) = g(x)(A)$.

2. The sequential execution of statements corresponds to the Kleisli product of single executions:

$$C[\beta_1; \beta_2] := C[\beta_2] \cdot C[\beta_1].$$

3. The assignment is given through the evaluation function:

$$C[a := E]g := E(a)(E) \cdot g.$$

4. The conditional statement permits operating on the true branch separately from the false branch:

$$C[\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}] := |\beta|_{\alpha = \text{tt}} + |\gamma|_{\alpha = \text{ff}},$$

provided $C[\beta] = |\beta|$ and $C[\gamma] = |\gamma|$. For the special case of the one-armed conditional statement we define

$$C[\text{if } \alpha \text{ then } \beta \text{ fi}]g := C[\text{if } \alpha \text{ then } \beta \text{ else skip fi}].$$

5. Modelling the transformation of the while- statement: If we are in state $g$, the resulting state after executing the statement not at all will be $\chi_{\{\alpha = \text{ff}\}}^g$. Suppose statement $\beta$ is associated with $|\beta|$, then executing the loop exactly $k$ times will result in the state $\chi_{\{\alpha = \text{ff}\}}^g \cdot |\beta|_{\alpha = \text{tt}}^k \cdot g$. Consequently we put

$$C[\text{while } \alpha \text{ do } \beta \text{ od}] := \rho_\alpha(|\beta|).$$
An easy induction on the length of a program $P$ yields

**Proposition 2.4.8** The semantic function $C[P]$ maps continuous stochastic relations to continuous stochastic relations. ⊥

An easy computation yields for the while-loop the equality one would intuitively expect.

**Lemma 2.4.9** $C[\text{while } \alpha \text{ do } \beta \text{ od}] = C[\text{if } \alpha \text{ then } \beta \text{ else } \beta; \text{while } \alpha \text{ do } \beta \text{ od fi}].$

**Proof** Let $T := C[\text{while } \alpha \text{ do } \beta \text{ od}]$, then

$$T = |\text{skip}|_{\alpha = \text{ff}} + \sum_{k=0}^{\infty} |\text{skip}|_{\alpha = \text{ff}} \cdot (|\beta|_{\alpha = \text{tt}})^{k+1}$$

$$= |\text{skip}|_{\alpha = \text{ff}} + \sum_{k=0}^{\infty} |\text{skip}|_{\alpha = \text{ff}} \cdot (|\beta|_{\alpha = \text{tt}})^{k} \cdot (|\beta| \cdot \chi^{\text{tt}}_{\alpha = \text{tt}}).$$

On the other hand, expanding $C[\text{if } \alpha \text{ then } \beta \text{ else } \beta; \text{while } \alpha \text{ do } \beta \text{ od fi}]$ we see that the latter expression equals

$$|\beta; \text{while } \alpha \text{ do } \beta \text{ od}|_{\alpha = \text{tt}} + |\text{skip}|_{\alpha = \text{ff}} = |\text{skip}|_{\alpha = \text{ff}} + T \cdot (|\beta| \cdot \chi^{\text{tt}}_{\alpha = \text{tt}})$$

$$= \sum_{k=0}^{\infty} |\text{skip}|_{\alpha = \text{ff}} \cdot (|\beta|_{\alpha = \text{tt}})^{k} \cdot (|\beta| \cdot \chi^{\text{tt}}_{\alpha = \text{tt}})$$

$$+ |\text{skip}|_{\alpha = \text{ff}},$$

establishing the claim. ⊥

### 2.4.4 A Partial Correctness Logic

We will give rules which permit statements of the form $f \{P\} g$ indicating that if execution of $P$ started in state $f$, and if $P$ terminates, then the computation will be in state $g$. Here states are continuous stochastic relations.

1. **skip** does not change much:

   $f \{\text{skip}\} f$

2. Compositionality is preserved:

   $$\frac{f \{\beta\} g \ g \{\gamma\} h}{f \{\beta; \gamma\} h}$$

3. The assignment is modelled through the evaluation $\mathcal{E}$

   $$g \circ \mathcal{E}(a)(\mathcal{E}) \{a := \mathcal{E}\} g$$

4. The predicate transformation through the conditional statement models the composition through both branches

   $$\chi^{\text{tt}}_{\alpha = \text{tt}} \cdot f \{\beta\} \chi^{\text{tt}}_{\alpha = \text{tt}} \cdot g \ \chi^{\text{tt}}_{\alpha = \text{ff}} \cdot f \{\gamma\} \chi^{\text{tt}}_{\alpha = \text{ff}} \cdot g$$

   $$f \{\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}\} g$$
5. The while-statement separates the first step from the rest of the iteration:

\[
\begin{align*}
\frac{h\{\beta\} \chi_{\alpha=tt}^* f \{\text{while } \alpha \text{ do } \beta \text{ od}\} g}{(\chi_{\alpha=tt}^* g + f \cdot h) \{\text{while } \alpha \text{ do } \beta \text{ od}\} g}
\end{align*}
\]

It would be of course desirable to have a simple rule for the while-statement, for example the simple and elegant rule working with invariants

\[
\frac{F \land \alpha \{\text{while } \alpha \text{ do } \beta \text{ od}\} F}{F \{\text{while } \alpha \text{ do } \beta \text{ od}\} F} \land \neg \alpha
\]

discussed in [65, Chapter 6.5]. Having a look, however, at the context in which consistency and completeness are discussed, it becomes obvious that this approach works only because the domain under consideration is flat. Consequently, the price for our more general and differently structured domain is inevitably a system with more complex rules.

### 2.4.5 Consistency and Completeness

As usual, a triplet \(\langle f, P, g \rangle\) is just \(f \{P\} g\), and a proof of triplet \(T_0, \ldots, T_\ell\) is a sequence of triplets where each triplet \(T_i\) follows either the previous one by one of the rules above, or is one of the axioms. Denote by \(P, f \vdash g\) that there is a proof for triplet \(\langle f, P, g \rangle\).

The following observation is quite immediate.

**Lemma 2.4.10** Let \(\beta, \gamma\) be program statements, and assume that \(\beta, f \vdash g\) iff \(C[\beta]g = f\) as well as \(\gamma, f \vdash g\) iff \(C[\gamma]g = f\) hold for all continuous stochastic relations \(f, g\). Then

1. if \(\alpha\) then \(\beta\) else \(\gamma\) fi, \(f \vdash g\) implies \(C[\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}]g = f\).
2. \(C[\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}]g = f\) implies if \(\alpha\) then \(\beta\) else \(\gamma\) fi, \(f \vdash g\).

**Proof** 1. To establish the first part, assume

\(f \{\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}\} g\),

thus both

\(\chi_{\alpha=tt}^* f \{\beta\} \chi_{\alpha=tt}^* g\) and \(\chi_{\alpha=ff}^* f \{\gamma\} \chi_{\alpha=ff}^* g\)

hold. From the assumption we conclude

\(C[\beta](\chi_{\alpha=tt}^* g) = \chi_{\alpha=tt}^* f\) and \(C[\gamma](\chi_{\alpha=ff}^* g) = \chi_{\alpha=ff}^* f\).

Since

\(|\beta|^* \chi_{\alpha=tt}^* f = |\beta|_{\alpha=tt}\)

and similarly for \(|\gamma|_{\alpha=ff}\), we obtain

\(f = |\beta|_{\alpha=tt}^* (g) + |\gamma|_{\alpha=tt}^* (g) = C[\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}]g = f\).

2. Assume for the second part that \(C[\text{if } \alpha \text{ then } \beta \text{ else } \gamma \text{ fi}]g = f\),
so that
\[ f = |\beta|_{\alpha=tt} \ast (g) + |\gamma|_{\alpha=tt} \ast (g), \]
equivalently,
\[ f = |\beta| \ast (\chi_{\{\alpha=tt\}} \ast g) + |\gamma| \ast (\chi_{\{\alpha=ff\}} \ast g). \]

We can find \( h_1, h_2 \) with
\[ h_1 = |\beta| \ast (\chi_{\{\alpha=tt\}} \ast g) \text{ and } h_2 = |\gamma| \ast (\chi_{\{\alpha=ff\}} \ast g) \]
so that
\[ f = h_1 + h_2 \]
(e.g. \( h_1 := C[if \alpha \text{ then } \beta \text{ else } \gamma fi](\chi_{\{\alpha=tt\}} \ast g) \))
would do, similar for \( h_2 \). The assumption implies that
\[ \beta, h_1 \vdash \chi_{\{\alpha=tt\}} \ast g, \text{ and } \gamma, h_2 \vdash \chi_{\{\alpha=ff\}} \ast g. \]
Hence we find \( h \) with
\[ h_1 = \chi_{\{\alpha=tt\}} \ast h \text{ and } h_2 = \chi_{\{\alpha=ff\}} \ast h. \]
Since \( h_1 + h_2 = f \), we infer
\[ h_1 = \chi_{\{\alpha=tt\}} \ast h = \chi_{\{\alpha=tt\}} \ast f, \]
similarly,
\[ h_2 = \chi_{\{\alpha=ff\}} \ast f. \]
But this implies
\[ \text{if } \alpha \text{ then } \beta \text{ else } \gamma fi, f \vdash g, \]
and we are done. \( \dashv \)

**Proposition 2.4.11** Given a program \( P \) and two continuous stochastic relations \( f \) and \( g \), the following statements are equivalent

1. \( P, f \vdash g \),
2. \( C[\![P]\!]g = f \).

**Proof**

1. (Consistency) We show that if \( P, f \vdash g \) then \( C[\![P]\!]g = f \) by induction on the length of program \( P \). We will consider in the inductive step the while-loop and refer for the conditional to Lemma 2.4.10. Suppose
\[ f \{\text{while } \alpha \text{ do } \beta \text{ od}\} g, \]
then this leads together with the induction hypothesis to the recursive equation
\[ f = (\chi_{\{\alpha=ff\}} \ast f + |\beta| \ast \chi_{\{\alpha=tt\}}) \ast g, \]
and since the loop is assumed to terminate, we obtain from Corollary 2.4.7 that
\[
f = (\rho_{(\alpha=tt)}(|\beta|))^\ast(g) = C[\text{while } \alpha \text{ do } \beta \text{ od}]g.
\]

2. (Completeness) We show that \(C[P]g = f\) implies \(P, f \vdash g\). This is also established by an inductive proof, and again we refer for the conditional statement to Lemma 2.4.10. For the while-statement to assertion follows immediately from Corollary 2.4.7, since
\[
C[\text{while } \alpha \text{ do } \beta \text{ od}]g = (\rho_{(\alpha=tt)}(|\beta|))^\ast(g).
\]

\[
\]

As an illustration, we get from Proposition 2.4.11 together with Lemma 2.4.9 e.g. the usual equivalence of the while-loop with a conditional that has a while inside.

**Corollary 2.4.12** Given continuous stochastic relations \(f, g\), the following statements are equivalent

1. while \(\alpha\) do \(\beta\) od, \(f \vdash g\)

2. if \(\alpha\) then \(\beta\) else \(\beta\); while \(\alpha\) do \(\beta\) od fi, \(f \vdash g\)

\[
\]

2.5 Bibliographic Notes

**Categorial Aspects Of Probabilities.** M. Giry’s paper [39] seems to be the first systematic investigation of categorial aspects of probability spaces on Polish and general measurable spaces (while at that time there was a substantial body of results on the connection between measure spaces on compact spaces and convex sets [81]); the main line of development pursued here is taken from that paper. P. Panangaden [72, 71] used Giry’s construction to elaborate on the analogy between set-theoretic and probabilistic relations; this line of argumentation has also been emphasized in the present exposition. Panangaden’s construction in [72] is slightly different from the one discussed here: he takes two measurable spaces \(X\) and \(Y\) and uses a stochastic transformation as a morphism between \(X\) and \(Y\); composition of morphisms is then given by the Kleisli product (in analogy to policies and randomized policies in stochastic dynamic optimization [80], one might call them randomized morphisms). In later papers, e.g. [20, 21], however, the model presented here has been used, albeit usually in a coalgebraic context, for modelling state transitions. One important field of investigations has been labelled Markov transition systems. They are constituted of a state space \(S\), a set \(A\) of actions, and for each action a probabilistic relation \(k_a : S \rightsquigarrow S\), so that \(k_a(s)(T)\) indicates the probability that the system’s state will be a member of Borel set \(T \subseteq S\) upon action \(a \in A\) in state \(s \in S\). Such a transition system serves as a probabilistic Kripke model for a very simple modal logic without negation that, as K. Larsen and A. Skou [59] have forcefully demonstrated, is useful in testing. Bisimilarity of systems with its intimate connections to testing is illustrated through the Hennessy-Milner Theorem and its ramifications. We will discuss this in chapter 5.

The categorial notions and constructions are taken from Mac Lane’s book [62, Chapter VI], where monads, the Kleisli construction and algebras are discussed in painstaking detail.
Architectural Issues. Since software architecture is a lively field of research in software engineering that has already entered most curricula in this area [75, 27], many different approaches on diverse levels can be found in the literature. Apart from formal approaches, architectural description languages formulate abstractions closer to implementations than, say, an approach resting on category theory; see [84] for a discussion and assessment of these languages. Other formalisms are used as well. The paper [64] by Medvidovic at al. investigates the suitability of the Unified Modelling Language for architectural descriptions. Probably more important, it discusses some desiderata for the language to be usable for architectural descriptions. Pipes and filters are targeted in the work reported in [1, 85]. The discussion centers around a formulation of this architecture through a denotational framework for developing formal models of architectural styles. It is based on the specification language Z. Arbab and Rutten present Reo, a calculus of component connectors based on coinduction [3]; coalgebras play a leading role in Barbosa’s work on components [8] as well. A formal calculus of connectors is given in [14], assigning connectors the rôle as mediators for the interaction between other computational components and connectors; formally, this model is an attempt to conciliate between categorical and the algebraic approaches to interaction. The FOCUS calculus developed by Broy and Stølen [13] is based on channels that are used by components to exchange information in terms of messages. Category theory is used in formalizations e.g. of architectures for mobile programs based on UNITY [96, 34]. Finally, Lajios [57] shows that additional assumptions are needed when decisions are to be modelled in an architecture (a decision might involve the selection for the next component to be visited); he models layered architectures using lextensive categories [16] and shows that tools gleaned from graph transformations are a most welcome addition to modelling architectural transformations.

Semantic Aspects. A development towards the semantics of probabilistic programs in terms of the Eilenberg-Moore algebras on probabilistic powerdomains is presented in [48]: Jones shows among others how programs can be understood as state transformers using upper and lower continuous functions with evaluations (for which an integration theory is developed); the underlying computational model is Moggi’s $\lambda_c$ calculus. An early proposal for using probabilities for modellng in automata theory can be found as an illustrating example in [4, Nos. (6), (13)]. The discussion assumes probabilities with finite support. Heckmann [43] discusses different approaches to probabilistic domains and addresses the question of the equivalence of different axiomatizations. His most powerful theory is called Multiple Choice with Divergence, and the axioms come very close to the axioms for positive convex sets — he even uses the notion of a formal linear combination $\sum_{i=1}^n p_i : x_i$. The theory is related to the probabilistic powerdomains investigated by Jones and Plotkin [48, 49]. No connection is drawn by Heckmann to probabilistic scenarios based on other than finite sets, and no relation to categorical constructions is attempted.

Using programs as transformers has been used for quite some time when the objects to be transformed may be cast as predicates in some logic, e.g. for Hoare triplets. The use of probabilities is not quite as common. Kozen uses in [55] a predicate transformer technique in his analysis of probabilistic programs; he demonstrated essentially how constructs for randomizing give rise to a transform on a Banach space of continuous...
functions. Kozen assumes essentially that the input follow a probability distribution, so that programs work as measure transforming devices. We go beyond this by assuming that each variable has its own probability distribution (so that the programs under consideration now transform families of distributions rather than single ones). Hence Kozen’s approach may be subsumed under the present one. It would be interesting to see if his functional approach does as well: if $g : X \sim X$ is a continuous stochastic relation, then $g$ spawns a linear operator $T_g : \mathcal{C}(X) \to \mathcal{C}(X)$ with $(T_g f)(x) := \int_X f \, dg(x)$. Here $\mathcal{C}(X)$ is as in section A.3.1 the Banach space of all bounded and continuous maps $f : X \to \mathbb{R}$ with the sup-norm $|| \cdot ||_\infty$. Since $g : X \to \mathcal{S}(X)$, we have

$$||T_g|| = \sup\{||T_g f||_\infty \mid ||f||_\infty \leq 1\} \leq 1,$$

thus $T_g$ is strongly continuous. This may then be applied: a while-loop is associated with the operator $T_{\rho_{(\alpha \Rightarrow \tau)}(\beta)}$, and studying the semantics of Ludwig then amounts to studying the behavior of linear operators on $\mathcal{C}(X)$ (see [74] for a similar approach for probabilistic concurrent constraint programming).

For deterministic programs, the present author showed e.g. in [24] how the notion of programs as measure transforms can be used for the average case analysis of sorting algorithms, in particular heapsort. Monniaux’s work, e.g. [68], uses probabilistic methods for the abstract interpretation of programs. He proposes among others an adjoint semantics of nondeterministic, probabilistic programs, and relates this to abstract interpretations of non-probabilistic programs. The methods employed come specifically from the duality theory of linear operators on integrable functions.
Chapter 3

Eilenberg-Moore Algebras for Stochastic Relations

Contents

3.1 Algebras for a Monad .............................................. 72
3.2 Characterization Through Equivalence Relations ............... 73
  3.2.1 Preparations .............................................. 74
  3.2.2 Positive Convex Partitions ............................... 75
  3.2.3 Smooth Relations ......................................... 78
3.3 Positive Convex Structures ..................................... 81
3.4 Algebras Through Positive Convex Structures ..................... 82
3.5 Examples .......................................................... 85
  3.5.1 Monad multiplication ...................................... 85
  3.5.2 The Finite Case ........................................... 86
  3.5.3 The Unit Interval ......................................... 87
  3.5.4 Barycenter ................................................. 87
3.6 The Left Adjoint .................................................. 89
3.7 Case Study: Derandomization .................................... 92
  3.7.1 Derandomizations as Algebras ............................ 92
  3.7.2 Derandomizing Ludwig .................................... 93
3.8 Bibliographic Notes ............................................. 94

It is shown in [62, Chapter VI] that the adjunction constructed from the Eilenberg-Moore algebras and the one constructed through the Kleisli category form in some sense the extreme points in a category of all adjunctions from which the given monad can be recovered. From this, the algebraic interest to identify these algebras is derived. The algebras for the power set monad (dubbed here the Manes monad) are identified in [62, Exercise VI.2.1], and we will identify the algebras for the sub-probability functor through smooth equivalence relations and through positive convex structures in this chapter. After some preliminary work that recalls the definition of an algebra, and that defines the mentioned structures, we will characterize these algebras first through the equivalence relations they induce on the set of sub-probability measures. This will be a vehicle for
an identification of these algebras without having to refer to the underlying probabilistic structure. This is done for the sub-probability functor and, with some small deviations, for the probability functor as well. We provide some examples to illustrate the algebras. Finally, the left adjoint of the forgetful functor that assigns each algebra the underlying Polish space is identified; it is just the functor that maps each Polish space to all its sub-probabilities (with the monad’s multiplication as the associated algebra).

The identification of algebras through convex structures is applied as a case study to the probabilistic semantics of Ludwig-programs. We know how a program transforms stochastic relations through the partial correctness semantics studied in section 2.4, and we show how this carries over to consistent decisions.

We work in this chapter in the category $c\text{Pol}$ of Polish spaces with continuous maps as morphisms. A possible and desirable extension to the discussion here would be identification of Eilenberg-Moore algebras for the sub-probability functor on analytic spaces with Borel measurable maps as morphisms.

### 3.1 Algebras for a Monad

The Kleisli construction that was introduced in section 2.1 and has been closer investigated in the previous chapter helps in constructing an adjunction from which the monad can be recovered. Technically this is done by constructing a functor from the category into the Kleisli category, from which an adjoined functor is easily determined. This observation was historically the original motivation for having a closer look at the Kleisli construction. Eilenberg-Moore algebras are another way of doing this.

Given a monad $\langle \mathcal{T}, e, m \rangle$ in a category $C$, a pair $\langle x, h \rangle$ consisting of an object $x$ and a morphism $h : \mathcal{T}x \to x$ in $C$ is called an Eilenberg-Moore algebra for the monad iff the following diagrams commute

\[
\begin{array}{c}
\mathcal{T}x \\ m_x \\
\downarrow h \\
\mathcal{T}x \\
\end{array}
\quad
\begin{array}{c}
x \\ x \downarrow \\
\mathcal{T}x \\
\end{array}
\quad
\begin{array}{c}
\mathcal{T}h \\ x \\
\downarrow h \\
x \\
\end{array}
\]

An algebra morphism $f : \langle x, h \rangle \to \langle x', h' \rangle$ between the algebras $\langle x, h \rangle$ and $\langle x', h' \rangle$ is a morphism $f : x \to x'$ which makes the diagram commute.

\[
\begin{array}{c}
\mathcal{T}x \\ h \\
\downarrow f \\
\mathcal{T}x' \\
\end{array}
\quad
\begin{array}{c}
x \\ f \\
\downarrow \\
x' \\
\end{array}
\]

Algebras together with their morphisms form a category $\text{Alg}_{\langle \mathcal{T}, e, m \rangle}$. We will usually omit the reference to the monad.
We mention as an illustration the algebras for the Manes monad \( \langle \mathcal{P} \mathcal{W}, e, m \rangle \) in the category Set of sets that is defined in section 2.1 very briefly. It is well known that the algebras for this monad may be identified with the complete sup-semi lattices \([62, \text{Exercise VI.2.1}]\). Doing this exercise is instructive for observing the components of a monad at work.

Assume first that \( \leq \) is a partial order on a set \( X \) that is sup-complete, so that \( \sup a \) exists for each \( a \subseteq X \). Define \( h(a) := \sup a \), then we have for each \( A \in \mathcal{P} \mathcal{W} \mathcal{W}(\mathcal{P} \mathcal{W}(X)) \) from the familiar properties of the supremum

\[
\sup(\bigcup A) = \sup \{ \sup a \mid a \in A \}.
\]

This translates into \((h \circ m_{\mathcal{W}})(A) = (h \circ \mathcal{P} \mathcal{W}(h))(A)\). Because \( x = \sup \{x\} \) holds for each \( x \in X \), we see that \( \langle X, h \rangle \) defines an algebra.

Assume on the other hand that \( \langle X, h \rangle \) is an algebra, and put for \( x, x' \in X \) \( x \leq x' \) iff \( h(\{x, x'\}) = x' \). This defines a partial order: reflexivity and antisymmetry are obvious, transitivity is seen as follows: assume \( x \leq x' \) and \( x' \leq x'' \), then

\[
\begin{align*}
\quad & h(\{x, x''\}) \\
= & \quad h(h(\{x\}), h(\{x', x''\})) \\
= & \quad (h \circ \mathcal{P} \mathcal{W}(h))(\{\{x\}, \{x', x''\}\}) \\
= & \quad (h \circ m_{\mathcal{W}})(\{\{x\}, \{x', x''\}\}) \\
= & \quad h(\{x, x', x''\}) \\
= & \quad (h \circ \mathcal{P} \mathcal{W}(h))(\{\{x, x'\}, \{x', x''\}\}) \\
= & \quad h(\{x', x''\}) \\
= & \quad x''.
\end{align*}
\]

It is clear from \( \{x\} \cup \emptyset = \{x\} \) for every \( x \in X \) that \( h(\emptyset) \) is the smallest element. Finally, is has to be shown that \( h(a) \) is the smallest upper bound for \( a \subseteq X \) in the order \( \leq \). We may assume that \( a \neq \emptyset \). Suppose that \( x \leq t \) holds for all \( x \in a \), then

\[
\begin{align*}
\quad & h(a \cup \{t\}) = h\left( \bigcup_{x \in a} \{x, t\} \right) \\
= & \quad (h \circ m_{\mathcal{W}})(\{\{x, t\} \mid x \in a\}) \\
= & \quad (h \circ \mathcal{P} \mathcal{W}(h))(\{\{x, t\} \mid x \in a\}) \\
= & \quad h(\{h(\{x, t\}) \mid x \in a\}) \\
= & \quad h(\{\{t\}\}) \\
= & \quad t.
\end{align*}
\]

Thus, if \( x \leq t \) for all \( x \in a \), hence \( h(a) \leq t \), thus \( h(a) \) is an upper bound to \( a \), and similarly, \( h(a) \) is the smallest upper bound. —

We will turn now to a characterization of the Eilenberg-Moore algebras for the Giry monad over some Polish space \( X \) that will be fixed throughout.

### 3.2 Characterization Through Equivalence Relations

We will show in this section that an algebra may be characterized in the way its fibres, i.e., the inverse images of points, partition the domain \( \mathcal{W}(X) \). The aspect that interests
here is that these partitions are positive convex and take closed values, they have an additional property due to continuity. This yields necessary and sufficient conditions for the characterization of partitions spawned by these algebras, a characterization of the morphisms in the category of all algebras is also derived.

3.2.1 Preparations

We need some elementary properties for later reference. They are collected in the next Lemma.

**Lemma 3.2.1**

1. Let $f : A \rightarrow B$ be a map between the Polish spaces $A$ and $B$, and let
   \[ \mu = \alpha_1 \cdot \delta_{a_1} + \ldots + \alpha_n \cdot \delta_{a_n} \]
   be the linear combination of Dirac measures for $a_1, \ldots, a_n \in A$ with positive convex $\langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega$. Then $\mathcal{S}(f)(\mu) = \alpha_1 \cdot \delta_{f(a_1)} + \ldots + \alpha_n \cdot \delta_{f(a_n)}$.

2. Let $\mu_1, \ldots, \mu_n$ be sub-probability measures on $X$, and let
   \[ M = \alpha_1 \cdot \delta_{\mu_1} + \ldots + \alpha_n \cdot \delta_{\mu_n} \]
   be the linear combination of the corresponding Dirac measures in $\mathcal{S}(\mathcal{S}(X))$ with positive convex coefficients $\langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega$. Then $m_X(M) = \alpha_1 \cdot \mu_1 + \ldots + \alpha_n \cdot \mu_n$.

**Proof** The first part follows directly from the observation that $\delta_x(f^{-1}[D]) = \delta_{f(x)}(D)$, and the second one is easily inferred from

\[ m_X(\delta_{\mu})(Q) = \int_{\mathcal{S}(X)} \delta_{\mu}(d\rho) = \mu(Q) \]

for each Borel subset $Q \subseteq X$, and from the linearity of the integral. \( \dashv \)

Both $e_X$ and $m_X$ are morphisms in $\text{cPol}$ for Polish $X$, as the following Lemma shows.

**Lemma 3.2.2** $e_X : X \rightarrow \mathcal{S}(X)$ and $m_X : \mathcal{S}(\mathcal{S}(X)) \rightarrow \mathcal{S}(X)$ are continuous.

**Proof**

1. Continuity of $e_X$ is clear, since $x_n \rightarrow x$ implies
   \[ \int_X f \, d e_X(x_n) = \int_X f \, d \delta_{x_n} = f(x_n) \rightarrow f(x) = \int_X f \, d e_X(x), \]
   whenever $f \in \mathcal{C}(X)$ is continuous and bounded. Thus $e_X(x_n) \rightarrow_w e_X(x)$.

2. Let $(M_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{S}(\mathcal{S}(X))$ with $M_n \rightarrow_w M_0$, then we get for $f \in \mathcal{C}(X)$ through the Change of Variable formula, and because
   \[ \mu \mapsto \int_X f \, d\mu \]
is a member of $C(S(X))$, the following chain
\[
\int_{\mathcal{S}(X)} f \, d\mu(X) = \int_{\mathcal{S}(X)} \left( \int_X f \, d\mu \right) \, d\mu(M_n) \\
\rightarrow \int_{\mathcal{S}(X)} \left( \int_X f \, d\mu \right) \, d\mu(M_0) \\
= \int_{\mathcal{S}(X)} f \, d\mu(M_0).
\]
Thus $m_X(M_n) \rightarrow_w m_X(M_0)$ is established, as desired. ⊣

We put
\[
\Omega := \{ \langle \alpha_1, \ldots, \alpha_k \rangle \mid k \in \mathbb{N}, \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i \leq 1 \}
\]
for the rest of the chapter, the elements of $\Omega$ being called positive convex tuples or simply positive convex.

### 3.2.2 Positive Convex Partitions

The natural approach is to think of these algebras in terms of an equivalence relation which may be thought to identify probability distributions, and to investigate either these relations or the partitions associated with them. These characterizations lead to the identification of the algebras as exactly the positive convex structures on their base space.

Assume that the pair $(X, h)$ is an algebra, and define for each $x \in X$
\[
G_h(x) := \{ \mu \in \mathcal{S}(X) \mid h(\mu) = x \} = h^{-1}([x])
\]
Then $G_h(x) \neq \emptyset$ for all $x \in X$ due to $h$ being onto. The algebra $h$ will be characterized through properties of the set-valued map $G_h$. We will need the weak inverse $\exists R$ for a set-valued map $R : X \to \mathcal{P}(Y) \setminus \{\emptyset\}$, see section A.2.3. If $Y$ is a topological space, if $R$ takes closed values, and if $\exists R(W)$ is compact in $X$ whenever $W \subseteq Y$ is compact, then $R$ is called $k$-upper-semicontinuous (abbreviated as k.u.s.c.). If $Y$ is compact, this is the usual notion of upper-semicontinuity known from topology.

The importance of being k.u.s.c. becomes clear at once from

**Lemma 3.2.3** Let $f : A \to B$ be a surjective map between the Polish spaces $A$ and $B$, and put $G_f(b) := f^{-1}([b])$ for $b \in B$. Then $f$ is continuous iff $G_f$ is k.u.s.c.

**Proof** A direct calculation for the weak inverse shows $\exists G_f(A_0) = f[A_0]$ for each subset $A_0 \subseteq A$. The assertion now follows from the well-known fact that a map between metric spaces is continuous iff it maps compact sets to compact sets. ⊣

Applying this observation to the set-valued map $G_h$, we obtain:

**Proposition 3.2.4** The set-valued map $x \mapsto G_h(x)$ has the following properties:

1. $\delta_x \in G_h(x)$ holds for each $x \in X$.
2. $\mathcal{G}_h := \{G_h(x) \mid x \in X\}$ is a partition of $\mathcal{S}(X)$ into closed and positive convex sets.
3. \( x \mapsto G_h(x) \) is k.u.s.c.

4. Let \( \sim_h \) be the equivalence relation on \( \mathcal{G}(X) \) induced by the partition \( G_h \). If \( \mu_i \sim_h \mu'_i \) (1 \( \leq i \leq n \)), then

\[
(\alpha_1 \cdot \mu_1 + \ldots + \alpha_n \cdot \mu_n) \sim_h (\alpha_1 \cdot \mu'_1 + \ldots + \alpha_n \cdot \mu'_n)
\]

for the positive convex coefficients \((\alpha_1, \ldots, \alpha_n) \in \Omega\).

**Proof** Because \( \{x\} \) is closed, and \( h \) is continuous, \( G_h(x) = h^{-1}\{x\} \) is a closed subset of \( \mathcal{G}(X) \). Because \( h \) is onto, every \( G_h \) takes non-empty values; it is clear that \( \{G_h(x) \mid x \in X\} \) forms a partition of \( \mathcal{G}(X) \). Because \( h \) is continuous, \( G_h \) is k.u.s.c. by Lemma 3.2.3. Positive convexity will follow immediately from part 4.

Assume that \( h(\mu_i) = h(\mu'_i) = x_i \) (1 \( \leq i \leq n \)), and observe that \( h(\delta_x) = x \) holds for all \( x \in X \). Using Lemma 3.2.1, we get:

\[
h(\alpha_1 \cdot \mu_1 + \ldots + \alpha_n \cdot \mu_n) = (h \circ m_X) (\alpha_1 \cdot \delta_{\mu_1} + \ldots + \alpha_n \cdot \delta_{\mu_n})
= (h \circ \mathcal{G}(h)) (\alpha_1 \cdot \delta_{\mu_1} + \ldots + \alpha_n \cdot \delta_{\mu_n})
= h (\alpha_1 \cdot \delta_{h(\mu_1)} + \ldots + \alpha_n \cdot \delta_{h(\mu_n)})
= h (\alpha_1 \cdot \delta_{x_1} + \ldots + \alpha_n \cdot \delta_{x_n})
\]

In a similar way, \( h(\alpha_1 \cdot \mu'_1 + \ldots + \alpha_n \cdot \mu'_n) = h (\alpha_1 \cdot \delta_{x_1} + \ldots + \alpha_n \cdot \delta_{x_n}) \) is obtained. This implies the assertion.

Thus \( G_h \) is invariant under taking positive convex combinations. It is a positive convex partition in the sense of the following definition.

**Definition 3.2.5** An equivalence relation \( \rho \) on \( \mathcal{G}(X) \) is said to be positive convex iff \( \mu_i \rho \mu_i' \) for 1 \( \leq i \leq n \) and \((\alpha_1, \ldots, \alpha_n) \in \Omega\) together imply

\[
(\alpha_1 \cdot \mu_1 + \ldots + \alpha_n \cdot \mu_n) \rho (\alpha_1 \cdot \mu'_1 + \ldots + \alpha_n \cdot \mu'_n)
\]

for each \( n \in \mathbb{N} \). A partition of \( \mathcal{G}(X) \) is called positive convex iff its associated equivalence relation is.

Note that the elements of a positive convex partition form positive convex sets. The converse to Proposition 3.2.4 characterizes algebras:

**Proposition 3.2.6** Assume \( G = \{G(x) \mid x \in X\} \) is a positive convex partition of \( \mathcal{G}(X) \) into closed sets which is indexed by \( X \) such that \( \delta_x \in G(x) \) for each \( x \in X \), and such that \( x \mapsto G(x) \) is k.u.s.c. Define \( h : \mathcal{G}(X) \to X \) through \( h(\mu) = x \) iff \( \mu \in G(x) \). Then \( (X, h) \) is an algebra for the Giry monad.

**Proof** 1. It is clear that \( h \) is well defined and surjective, and that

\[
\exists \mathcal{G}(F) = h[F]
\]

holds for each subset \( F \subseteq \mathcal{G}(X) \). Thus \( h[K] \) is compact whenever \( K \) is compact, because \( G \) is k.u.s.c. Hence \( h \) is continuous by Lemma 3.2.3.
2. An easy induction establishes that \( h \) respects positive convex combinations: if \( h(\mu_i) = h(\mu'_i) \) for \( i = 1, \ldots, n \), and if \( \alpha_1, \ldots, \alpha_n \) are positive convex coefficients, then

\[
h\left( \sum_{i=1}^{n} \alpha_i \cdot \mu_i \right) = h\left( \sum_{i=1}^{n} \alpha_i \cdot \mu'_i \right).
\]

We claim that

\[
(h \circ \mathfrak{m}_X)(M) = (h \circ \mathfrak{g}(h))(M)
\]

holds for each discrete \( M \in \mathfrak{g}(\mathfrak{S}(X)) \). In fact, let

\[
M = \sum_{i=1}^{n} \alpha_i \cdot \delta_{\mu_i}
\]

be such a discrete measure, then Lemma 3.2.1 implies that

\[
\mathfrak{m}_X(M) = \sum_{i=1}^{n} \alpha_i \cdot \mu_i,
\]

thus

\[
(h \circ \mathfrak{m}_X)(M) = h\left( \sum_{i=1}^{n} \alpha_i \cdot \mu_i \right) = h\left( \sum_{i=1}^{n} \alpha_i \cdot \delta_{h(\mu_i)} \right) = (h \circ \mathfrak{g}(h))(M),
\]

because we know also from Lemma 3.2.1 that

\[
\mathfrak{g}(h)(M) = \sum_{i=1}^{n} \alpha_i \cdot \delta_{h(\mu_i)}
\]

holds.

3. Since the discrete measures are dense in the weak topology (see section A.3.1), we find for \( M_0 \in \mathfrak{g}(\mathfrak{S}(X)) \) a sequence \( (M_n)_{n \in \mathbb{N}} \) of discrete measures \( M_n \) with \( M_n \xrightarrow{w} M_0 \). Consequently, we get from the continuity of both \( h \) and \( \mathfrak{m}_X \) (Lemma 3.2.2) together with the continuity of \( \mathfrak{g}(h) \)

\[
(h \circ \mathfrak{m}_X)(M_0) = \lim_{n \to \infty} (h \circ \mathfrak{m}_X)(M_n) = \lim_{n \to \infty} (h \circ \mathfrak{g}(h))(M_n) = (h \circ \mathfrak{g}(h))(M_0).
\]

This proves the claim. \( \dashv \)

We have established

**Proposition 3.2.7** The algebras \( \langle X, h \rangle \) for the Giry monad for Polish spaces \( X \) are exactly the positive convex k.u.s.c. partitions \( \{ G(x) \mid x \in X \} \) into closed subsets of \( \mathfrak{S}(X) \) such that \( \delta_x \in G(x) \) for all \( x \in X \). \( \dashv \)

We characterize the category \( \text{Alg} \) of all algebras for the Giry monad. To this end we package the properties of partitions representing algebras into the notion of a G-partition. They will form the objects of category \( \text{GPart} \).

**Definition 3.2.8** \( \mathcal{G} \) is called a G-partition for \( X \) iff

1. \( \mathcal{G} = \{ G(x) \mid x \in X \} \) is a positive convex partition for \( \mathfrak{S}(X) \) into closed sets indexed by \( X \),
2. $\delta_x \in G(x)$ holds for all $x \in X$,

3. the set-valued map $x \mapsto G(x)$ is k.u.s.c.

Define the objects of category $G\text{Part}$ as pairs $\langle X, G \rangle$ where $X$ is a Polish space, and $G$ is a $G$-partition for $X$. A morphism $f$ between $G$ and $G'$ will map elements of $G(x)$ to $G'(f(x))$ through its associated map $\mathcal{G}(f)$. Thus an element $\mu \in G(x)$ will correspond to an element $\mathcal{G}(f)(\mu) \in G'(f(x))$.

**Definition 3.2.9** A morphism for $G\text{Part}$ $f : \langle X, G \rangle \rightarrow \langle X', G' \rangle$ is a continuous map $f : X \rightarrow X'$ such that

$$G(x) \subseteq \mathcal{G}(f)^{-1}\left[G'(f(x))\right]$$

holds for each $x \in X$.

Define the functor $F : \text{Alg} \rightarrow G\text{Part}$ by associating with each algebra $\langle X, h \rangle$ its Giry partition $F(X, h)$ according to Proposition 3.2.7. Assume that $f : \langle X, h \rangle \rightarrow \langle X', h' \rangle$ is a morphism in $\text{Alg}$, and let $G = \{G(x) \mid x \in X\}$ resp. $G' = \{G'(x') \mid x' \in X'\}$ be the corresponding partitions. Then the properties of an algebra morphism yield

$$\mu \in \mathcal{G}(f)^{-1}\left[G'(f(x))\right] \iff \mathcal{G}(f)(\mu) \in G'(f(x))$$

$$\iff (h' \circ \mathcal{G}(f))(\mu) = f(x)$$

$$\iff (f \circ h)(\mu) = f(x).$$

Thus $\mu \in \mathcal{G}(f)^{-1}\left[G'(f(x))\right]$, provided $\mu \in G(x)$. Hence $f$ is a morphism in $G\text{Part}$ between $F(X, h)$ and $F(X', h')$. Conversely, let $f : \langle X, G \rangle \rightarrow \langle X', G' \rangle$ be a morphism in $G\text{Part}$ with $\langle X, G \rangle = F(X, h)$ and $\langle X', G' \rangle = F(X', h')$. Then

$$h(\mu) = x \iff \mu \in G(x)$$

$$\Rightarrow \mathcal{G}(f)(\mu) \in G'(f(x))$$

$$\iff h'(\mathcal{G}(f)(\mu)) = f(x),$$

thus $h' \circ \mathcal{G}(f) = f \circ h$ is inferred. Hence $f$ constitutes a morphism in category $\text{Alg}$. Summarizing, we have shown

**Proposition 3.2.10** The category $\text{Alg}$ of algebras for the Giry monad is isomorphic to the category $G\text{Part}$ of $G$-partitions. ☐

### 3.2.3 Smooth Relations

The characterization of algebras so far encoded the crucial properties into a partition of $\mathcal{G}(X)$, thus indirectly into an equivalence relation on that space. We can move directly to a particular class of these relations when looking at an alternative characterization of the algebras through smooth equivalence relations. In contrast to the characterization in 3.2 that started from the fibres $h^{-1}\{x\}$ we study here the kernel of $h$, i.e. the set

$$\ker(h) := \{\langle \mu, \mu' \rangle \mid h(\mu) = h(\mu')\}.$$ 

We characterize then algebras in terms of the kernel for the associated map, and we indicate how an algebra may be constructed from such a partition. The characterization
is interesting in its own right and permits another characterization of morphisms for algebras, but it will also help in giving an intrinsic characterization of algebras in terms of positive convex structures.

**Definition 3.2.11** An equivalence relation $\rho$ on a Polish space $A$ is called smooth iff there exists a Polish space $B$ and a Borel measurable map $f : A \to B$ such that $\rho = \ker(f)$.

Smooth equivalence relations are a helpful tool in the theory of Borel sets [88, 51]. They will turn out to be most interesting in the investigation of stochastic relations as well, so we will return to them later in chapter 5 and study them in greater detail there. Despite the later and more systematic treatment of the topic, we will state some properties that will be needed here immediately.

Denote for an equivalence relation $\rho$ on $A$ by $A/\rho$ the factor space, i.e., the set of all equivalence classes $[a]_\rho$, and by

$$\eta_\rho : A \to A/\rho$$

the canonical projection. For the Polish space $A$ with topology $T$ let $T/\rho$ be the final topology on $A/\rho$ with respect to the given topology $T$ and $\eta_\rho$, i.e., the largest topology $T'$ on $A/\rho$ which makes $\eta_\rho : T-T'$-continuous. Clearly a map $g : A/\rho \to B$ for a topological space $(B, S)$ is $T/\rho-S$-continuous iff $g \circ \eta_\rho : A \to B$ is $T-S$-continuous. We will need this observation in the proof of Proposition 3.2.12.

Now let $\langle X, h \rangle$ be an algebra for the Giry monad. Obviously $\rho_h := \ker(h)$ defines a smooth equivalence relation $\rho_h$ on the Polish space $\mathcal{G}(X)$. Its properties are summarized in

**Proposition 3.2.12** The equivalence relation $\rho_h$ is positive convex, each equivalence class $[\mu]_{\rho_h}$ is closed and positive convex, and the factor space $\mathcal{G}(X)/\rho_h$ is homeomorphic to $X$ when the former is endowed with the topology $W/\rho_h$, $W$ being the topology of weak convergence on $\mathcal{G}(X)$.

**Proof** 1. Positive convexity of $\rho_h$ follows from the properties of $h$ exactly as in the proof of Proposition 3.2.4. Positive convexity of the classes is inferred from this as well. Continuity of $h$ implies that the classes are closed sets.

2. Define $\chi_h([\mu]_{\rho_h}) := h(\mu)$ for $\mu \in \mathcal{G}(X)$. Then $\chi_h : X/\rho_h \to X$ is well defined and a bijection. Let $G \subseteq X$ be an open set, then

$$\eta_{\rho_h}^{-1} [\chi_h^{-1} [G]] = h^{-1} [G].$$

Because $W/\rho_h$ is the largest topology on $\mathcal{G}(X)/\rho_h$ that renders $\eta_{\rho_h}$ continuous, and because $h^{-1} [G] \subseteq \mathcal{G}(X)$ is open by assumption, we infer that $\chi_h^{-1} [G]$ is $W/\rho_h$-open. Thus $\chi_h$ is continuous. On the other hand, if $(x_n)_{n \in \mathbb{N}}$ is a sequence in $X$ converging to $x_0 \in X$, then $\delta_{x_n} \to_w \delta_{x_0}$ in $\mathcal{G}(X)$, thus $[\delta_{x_n}]_{\rho_h} \to [\delta_{x_0}]_{\rho_h}$ in $W/\rho_h$ by construction. Consequently $\chi_h^{-1}$ is also continuous. \(\square\)

Thus each algebra induces a G-triplet consisting of its kernel and a homeomorphism:

**Definition 3.2.13** A G-triplet $\langle X, \rho, \chi \rangle$ is a Polish space $X$ with a smooth and positive convex equivalence relation $\rho$ on $\mathcal{G}(X)$ such that $\chi : \mathcal{G}(X)/\rho \to X$ is a homeomorphism with $\chi([\delta_x]_{\rho}) = x$ for all $x \in X$. Here $\mathcal{G}(X)/\rho$ carries the final topology with respect to the weak topology on $\mathcal{G}(X)$ and $\eta_{\rho}$. 

79
Now assume that a G-triplet \( \langle X, \rho, \chi \rangle \) is given. Define \( h(\mu) := \chi([\mu]_\rho) \) for \( \mu \in \mathcal{G}(X) \). Then \( \langle X, h \rangle \) is an algebra for the Giry monad: \( h(\delta_x) = x \) follows from the assumption, and because \( h = \chi \circ \eta_\rho \), the map \( h \) is continuous. An argument very similar to that used in the proof of Proposition 3.2.4 shows that \( h \circ m_X = h \circ \mathcal{G}(h) \) holds; this is so since \( \rho \) is assumed to be positive convex.

**Definition 3.2.14** The continuous map \( f : X \to X' \) between the Polish spaces \( X \) and \( X' \) constitutes a G-triplet morphism \( f : \langle X, \rho, \chi \rangle \to \langle X', \rho', \chi' \rangle \) if these conditions hold:

1. \( \mu \rho \mu' \) implies \( \mathcal{G}(f)(\mu) \rho' \mathcal{G}(f)(\mu') \),
2. the diagram

\[
\begin{array}{ccc}
\mathcal{G}(X)/\rho & \xrightarrow{\mathcal{G}(f)_{\rho,\rho'}} & \mathcal{G}(X')/\rho' \\
\chi \downarrow & & \downarrow \chi' \\
X & \xrightarrow{f} & X'
\end{array}
\]

commutes, where

\[
\mathcal{G}(f)_{\rho,\rho'}([\mu]_\rho) := [\mathcal{G}(f)(\mu)]_{\rho'}
\]

G-triplets with their morphisms form a category \( \text{GTrip} \).

**Lemma 3.2.15** Each algebra morphism \( f : \langle X, h \rangle \to \langle X', h' \rangle \) induces a G-triplet morphism \( f : \langle X, \rho_h, \chi_h \rangle \to \langle X', \rho_{h'}, \chi_{h'} \rangle \).

**Proof** 1. It is an easy calculation to show that \( \mathcal{G}(f)(\mu) \rho_{h'} \mathcal{G}(f)(\mu) \) holds, provided \( \mu \rho_h \mu' \). This is so because \( f \) is a morphism for the algebras.
2. Since for each \( \mu \in \mathcal{G}(X) \) there exists \( x \in X \) such that \( [\mu]_{\rho_h} = [\delta_x]_{\rho_h} \) (in fact, \( h(\mu) \) would do, because \( h(\mu) = h(\delta_{h(\mu)}) \), as shown above), it is enough to demonstrate that

\[
\chi_{h'}(\mathcal{G}(f)_{\rho_h,\rho_{h'}}([\delta_x]_{\rho_h})) = f(\chi_h([\delta_x]_{\rho_h}))
\]

is true for each \( x \in X \). Because \( \mathcal{G}(f)(\delta_x) = \delta_{f(x)} \), a little computation shows that both sides of the above equation boil down to \( f(x) \). \( \dashv \)

The morphisms between G-triplets are just the morphisms between algebras (when we forget that these games play in different categories).

**Proposition 3.2.16** Let \( f : \langle X, \rho, \chi \rangle \to \langle X', \rho', \chi' \rangle \) be a morphism between G-triplets, and let \( \langle X, h \rangle \) resp. \( \langle X', h' \rangle \) be the associated algebras. Then \( f : \langle X, h \rangle \to \langle X', h' \rangle \) is an algebra morphism.

**Proof** Given \( \mu \in \mathcal{G}(X) \) we have to show that \( (f \circ h)(\mu) \) equals \( (h' \circ \mathcal{G}(f))(\mu) \). Since \( h(\mu) = \chi([\mu]_\rho) \), we obtain

\[
(f \circ h)(\mu) = f(\chi([\mu]_\rho)) = \chi'(\mathcal{G}(f)_{\rho,\rho'}([\mu]_\rho)) = \chi'(\mathcal{G}(f)(\mu)_{\rho'}) = (h' \circ \mathcal{G}(f))(\mu).
\]
Putting all these constructions with their properties together, we obtain as a second characterization.

**Proposition 3.2.17** The category $\text{Alg}$ of algebras for the Giry monad is isomorphic to the category $\text{GTrips}$ of G-triplets.

Albeit being treated similarly, the probabilistic case requires a separate discussion. We define an equivalence relation $\rho$ on $\mathcal{P}(X)$ to be convex iff for each $n \in \mathbb{N}$ the conditions $\mu_i \rho \mu_i'$ for $1 \leq i \leq n$ and $(\alpha_1, \ldots, \alpha_n) \in \Omega_c$ together imply $(\sum_{i=1}^n \alpha_i \cdot \mu_i) \rho (\sum_{i=1}^n \alpha_i \cdot \mu_i')$, where

$$\Omega_c := \{(\alpha_1, \ldots, \alpha_k) | \alpha_i \geq 0, \alpha_1 + \cdots + \alpha_k = 1\}$$

are all convex coefficients. Then the $\rho$-classes form convex subsets of $\mathcal{P}(X)$. We introduce PG-triplets $\langle X, \rho, \chi \rangle$ for a Polish space $X$, a smooth convex equivalence relation $\rho$ and a homeomorphism $\chi : \mathcal{P}(X)/\rho \to X$ with $\chi(\delta_{x\rho}) = x$ for all $x \in X$. A continuous map $f : X \to X'$ then is a PG-triplet morphism $\langle X, \rho, \chi \rangle \to \langle X', \rho', \chi' \rangle$ iff

1. $\mu \rho \mu' \Rightarrow \mathcal{P}(f)(\mu) = \mathcal{P}(f)(\mu')$,
2. $\chi' \circ \mathcal{P}(f)_{\rho,\rho'} = f \circ \chi$.

Here $\mathcal{P}(f)_{\rho,\rho'}$ is defined in analogy to $\mathcal{S}(f)_{\rho,\rho'}$ in Definition 3.2.14 as

$$\mathcal{P}(f)_{\rho,\rho'}([\mu]_\rho) := [\mathcal{P}(f)(\mu)]_{\rho'}.$$ 

We see then that each algebra morphism $f : \langle X, h \rangle \to \langle X', h' \rangle$ induces a PG-triplet morphism $f : \langle X, \rho_h, \chi_h \rangle \to \langle X', \rho_{h'}, \chi_{h'} \rangle$, and vice versa. The reader is invited to fill in the details.

Summarizing, this yields:

**Proposition 3.2.18** The category of algebras for the Giry monad for the probability functor is isomorphic to the full subcategory of G-triplets $\langle X, \rho, \chi \rangle$ with a smooth and convex equivalence relation such that $\chi : \mathcal{P}(X)/\rho \to X$ is a homeomorphism.

This prepares us for another and more self-contained characterization of algebras. We will show now that $\text{StrConv}$ is isomorphic to $\text{Alg}$.

### 3.3 Positive Convex Structures

Suppose the Polish space $X$ is embedded into a vector space $V$ over the reals as a positive convex structure. This means that, if $x_1, \ldots, x_k \in X$, $(\alpha_1, \ldots, \alpha_k) \in \Omega$, then $\sum_{i=1}^k \alpha_i \cdot x_i \in X$. In addition, forming positive convex combinations should be compatible with the topological structure on $X$, so it should be continuous. This means of course that $x_{i,n} \to x_{i,0}$ and $\alpha_n \to \alpha_0$ with $\alpha_0, \alpha_n \in \Omega$ together imply $\sum_{i=1}^k \alpha_{i,n} \cdot x_{i,n} \to \sum_{i=1}^k \alpha_{i,0} \cdot x_{i,0}$. These requirements are quite comparable to those for a topological vector space, postulating continuity of addition and scalar multiplication.

These observations meet the intuition about positive convexity, but it has the drawback that we have to look for the vector space $V$ into which $X$ to embed. It has the additional shortcoming that once we did identify $V$, the positive convex structure on $X$ is
fixed through the vector space, but we will see soon that we need some flexibility. Consequently, we propose an abstract description of positive convexity, much in the spirit of Pumplün’s approach [76]. Thus the essential properties (for us, that is) of positive convexity are described intrinsically for \( X \) without having to resort to a vector space. This leads to the definition of a positive convex structure.

**Definition 3.3.1** A positive convex structure \( P \) on the Polish space \( X \) has for each \( \alpha = \langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega \) a continuous map \( \alpha_P : X^n \to X \) which we write as

\[
\alpha_P(x_1, \ldots, x_n) = \sum_{1 \leq i \leq n} \alpha_i \cdot x_i,
\]

such that

1. \( \sum_{1 \leq i \leq n} \delta_{i,k} \cdot x_i = x_k \), where \( \delta_{i,j} \) is Kronecker’s \( \delta \) (thus \( \delta_{i,i} = 1 \) if \( i = j \), and \( \delta_{i,j} = 0 \), otherwise),
2. the identity

\[
\sum_{1 \leq i \leq n} \alpha_i \cdot \left( \sum_{1 \leq k \leq m} \beta_{i,k} \cdot x_k \right) = \sum_{1 \leq k \leq m} \left( \sum_{1 \leq i \leq n} \alpha_i \beta_{i,k} \right) \cdot x_k
\]

holds whenever \( \langle \alpha_1, \ldots, \alpha_n \rangle, \langle \beta_{i,k}, \ldots, \beta_{i,k} \rangle \in \Omega, 1 \leq i \leq n \).

Thus we will use freely the notation from vector spaces, omitting in particular the explicit reference to the structure whenever possible. Hence simple addition \( \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 \) will be written rather than \( \sum_{1 \leq i \leq 2} \alpha_i \cdot x_i \), with the understanding that it refers to a fixed positive convex structure \( P \) on \( X \).

It can be shown [76] that for a positive convex structure the usual rules for manipulating sums in vector spaces apply, e.g. \( 1 \cdot x = x, \sum_{i=1}^n \alpha_i \cdot x_i = \sum_{i=1}^n \alpha_i \cdot x_i \), or the law of associativity, \( (\alpha_1 \cdot x_1 + \alpha_2 \cdot x_2) + \alpha_3 \cdot x_3 = \alpha_1 \cdot x_1 + (\alpha_2 \cdot x_2 + \alpha_3 \cdot x_3) \). Nevertheless, care should be observed, for of course not all rules apply: we cannot in general conclude \( x = x' \) from \( \alpha \cdot x = \alpha \cdot x' \), even if \( \alpha \neq 0 \).

A morphism \( \theta : \langle X_1, P_1 \rangle \to \langle X_2, P_2 \rangle \) between continuous positive convex structures is a continuous map \( \theta : X_1 \to X_2 \) such that

\[
\theta \left( \sum_{1 \leq i \leq n} \alpha_i \cdot x_i \right) = \sum_{1 \leq i \leq n} \alpha_i \cdot \theta(x_i)
\]

holds for \( x_1, \ldots, x_n \in X \) and \( \langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega \). In analogy to linear algebra, \( \theta \) will be called an affine map. Positive convex structures with their morphisms form a category \( \text{StrConv} \).

### 3.4 Algebras Through Positive Convex Structures

The algebras are also described without having to resort to \( \mathcal{S}(X) \). This is done through an intrinsic characterization using positive convex structures with affine maps. This characterization is comparable to the one given by Manes for the power set monad (which also does not resort explicitly to the underlying monad or its functor), see page 73.
Lemma 3.4.1 Given an algebra \((X, h)\), define for \(x_1, \ldots, x_n \in X\) and the positive convex coefficients \((\alpha_1, \ldots, \alpha_n) \in \Omega\)

\[
\sum_{i=1}^{n} \alpha_i \cdot x_i := h(\sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i}).
\]

This defines a positive convex structure on \(X\).

Proof 1. Because

\[
h\left(\sum_{i=1}^{n} \delta_{i,j} \cdot \delta_{x_i}\right) = h(\delta_{x_j}) = x_j,
\]

property 1 in Definition 3.3.1 is satisfied.

2. Proving property 2, we resort to the properties of an algebra and a monad:

\[
\sum_{i=1}^{n} \alpha_i \cdot \left(\sum_{k=1}^{m} \beta_{i,k} \cdot x_k\right) = h\left(\sum_{i=1}^{n} \alpha_i \cdot \delta_{\sum_{k=1}^{m} \beta_{i,k} \cdot x_k}\right) = h\left(\sum_{i=1}^{n} \alpha_i \cdot \delta_{h(\sum_{k=1}^{m} \beta_{i,k} \cdot x_k)}\right) = h\left(\sum_{i=1}^{n} \alpha_i \cdot \delta_{h(\sum_{k=1}^{m} \beta_{i,k} \cdot x_k)}\right) = (h \circ \delta_{\delta_{\sum_{k=1}^{m} \beta_{i,k} \cdot x_k}})\left(\sum_{i=1}^{n} \alpha_i \cdot \delta_{\sum_{k=1}^{m} \beta_{i,k} \cdot x_k}\right) = h\left(\sum_{i=1}^{n} \alpha_i \cdot \delta_{m_X(\sum_{k=1}^{m} \beta_{i,k} \cdot x_k)}\right) = h\left(\sum_{i=1}^{n} \alpha_i \cdot \left(\sum_{k=1}^{m} \beta_{i,k} \cdot x_k\right)\right) = h\left(\sum_{k=1}^{m} \left(\sum_{i=1}^{n} \alpha_i \cdot \beta_{i,k}\right) \delta_{x_k}\right) = \sum_{k=1}^{m} \left(\sum_{i=1}^{n} \alpha_i \cdot \beta_{i,k}\right) x_k.
\]

The equations (3.1) and (3.2) reflect the definition of the structure, equation (3.3) applies \(\delta_{h(\tau)} = \delta_{\delta_{\sum_{k=1}^{m} \beta_{i,k} \cdot x_k}}\), equation (3.4) uses the linearity of \(\delta_{\delta_{\sum_{k=1}^{m} \beta_{i,k} \cdot x_k}}\) according to Lemma 3.2.1, equation (3.5) is due to \(h\) being an algebra. Winding down, equation (3.6) uses Lemma 3.2.1 again, this time for \(m_X\), equation (3.7) uses that \(m_X \circ \delta_a = \tau\), equation (3.8) is just rearranging terms, and equation (3.9) is the definition again. \(\square\)
Let conversely such a positive convex structure be given. We show that we can define a G-triplet from it. Let

\[ T_X := \{ \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i} \mid n \in \mathbb{N}, x_1, \ldots, x_n \in X, \langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega \}, \]

then \( T_X \) is dense in \( \mathcal{S}(X) \). Put

\[ h_0 \left( \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i} \right) := \sum_{i=1}^{n} \alpha_i \cdot x_i, \]

then \( h_0 : T_X \to X \) is well defined. This is so since

\[ \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i} = \sum_{j=1}^{m} \alpha_j' \cdot \delta_{x_j'} \]

implies that

\[ \sum_{i=1, \alpha_i \neq 0}^{n} \alpha_i \cdot \delta_{x_i} = \sum_{j=1, \alpha_j' \neq 0}^{m} \alpha_j' \cdot \delta_{x_j'}, \]

hence given \( i \) with \( \alpha_i \neq 0 \) there exists \( j \) with \( \alpha_j' \neq 0 \) such that \( x_i = x_j' \) and vice versa. Consequently,

\[ \sum_{i=1}^{n} \alpha_i \cdot x_i = \sum_{i=1, \alpha_i \neq 0}^{n} \alpha_i \cdot x_i = \sum_{j=1, \alpha_j' \neq 0}^{m} \alpha_j' \cdot x_j' = \sum_{j=1}^{m} \alpha_j' \cdot x_j' \]

is inferred from the properties of positive convex structures.

The map \( h_0 \) is uniformly continuous, because

\[ d \left( h_0 \left( \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i} \right), h_0 \left( \sum_{j=1}^{m} d_j \cdot \delta_{y_j} \right) \right) \leq d_P \left( \sum_{i=1}^{n} \alpha_i \cdot \delta_{x_i}, \sum_{j=1}^{m} d_j \cdot \delta_{y_j} \right), \]

\( d_P \) denoting the Prohorov metric, see section A.3. We need uniform continuity here, because it is well known that otherwise a unique, continuous extension from the dense subset of discrete measures to the set of all measures cannot be guaranteed.

Define \( \rho_0 \) as the kernel of \( h_0 \), then \( \rho_0 \) is a smooth equivalence relation on \( T_X \), and it is not difficult to see that the set of topological closures

\[ \{ \text{cl} \left( [t]_{\rho_0} \right) \mid t \in T_X \} \]

forms a partition of \( \mathcal{S}(X) \) through the following arguments:

1. the closures of different equivalence classes are disjoint,
2. given \( \mu \in \mathcal{S}(X) \), one can find a sequence \( (t_n)_{n \in \mathbb{N}} \) in \( T_X \) with \( t_n \to_w \mu \). Since \( X \) is Polish, in particular complete, the sequence \( (h_0(t_n))_{n \in \mathbb{N}} \) converges to some \( t_0 \), and because \( h_0 \) is uniformly continuous, one concludes that \( \mu \in \text{cl} \left( [t_0]_{\rho_0} \right) \). Thus each member of \( \mathcal{S}(X) \) is in some class.
This yields an equivalence relation \( \rho \) on \( \mathcal{S}(X) \). Uniform continuity of \( h_0 \) gives a unique continuous extension \( h \) of \( h_0 \) to \( \mathcal{S}(X) \), thus \( \rho \) equals the kernel of \( h \), hence \( \rho \) is a smooth equivalence relation, and it is evidently positive convex. Defining on \( \mathcal{S}(X)/\rho \) the metric
\[
D([\mu_1]_\rho, [\mu_2]_\rho) := d(h(\mu_1), h(\mu_2)),
\]
it is rather immediate that the metric space \( (\mathcal{S}(X)/\rho, D) \) is homeomorphic to \( X \) with metric \( d \), and that the topology induced by the metric is just the final topology with respect to the weak topology on \( \mathcal{S}(X) \) and \( \eta_\rho \).

It is clear that each affine and continuous map between positive convex structures gives rise to a morphisms between the corresponding G-triplets, and vice versa. Thus we have established:

**Proposition 3.4.2** The category of \( \text{Alg} \) of algebras for the Giry monad is isomorphic to the category \( \text{StrConv} \) of positive convex structures with continuous affine maps as morphisms. \( \dashv \)

For the probability functor we again mirror the development, but this time we need not go into details. We obtain eventually this characterization for the category \( \text{pAlg} \) of algebras for the Giry monad, when restricted to the probability functor (with the obvious necessary adjustments made for morphisms):

**Proposition 3.4.3** The category of algebras for the Giry monad for the probability functor is isomorphic to the full subcategory of continuous convex structures. \( \dashv \)

For a partial history of this result see the bibliographic notes at the end of this chapter.

### 3.5 Examples

We illustrate the concept and propose some examples by looking at some well-known situations. Most of this section is not really new, probably apart from the proposed point of view. We first show that the monad carries for each Polish space an instance of an algebra with it. Then we prove that in the finite case an algebra exists only in the case of a singleton set. Finally a geometrically oriented example is discussed by investigating the barycenter of a probability on a compact and convex subset of \( \mathbb{R}^n \).

In each case the geometry of the underlying space imposes a natural positive convex structure, and it invites itself to compare this structure with the one that can be constructed through the algebra. It turns out in each of these cases that the convex structure associated with the algebra is the natural one.

#### 3.5.1 Monad multiplication

We show that \( \langle \mathcal{S}(X), m_X \rangle \) is an algebra whenever \( X \) is a Polish space. This is not only interesting in its own right, it shows moreover that each Polish space is associated in a natural fashion with a strongly convex structure. This association entails actually more than meets the eye: we will show in section 3.6 that \( X \mapsto \langle \mathcal{S}(X), m_X \rangle \) is the object part of the left adjoint to the forgetful functor \( \text{Alg} \to \text{Pol} \).
Proposition 3.5.1 The pair \( (\mathcal{S}(X), m_X) \) is an algebra for each Polish space \( X \).

Proof We know from Lemma 3.2.2 that \( m_X : \mathcal{S}(\mathcal{S}(X)) \to \mathcal{S}(X) \) is continuous. Because \( \langle \mathcal{S}, \varepsilon, m \rangle \) is a monad, the natural transformation \( m : \mathcal{S}^2 \to \mathcal{S} \) satisfies \( m \circ \mathcal{S} = m \circ m \mathcal{S} \) in the category of functors with natural transformations as morphisms, see the diagram at the end of section 2.2. Since \( \mathcal{S}(m_X) = \mathcal{S}(m_S(X)) \) and \( (m \circ \mathcal{S})_X = m \mathcal{S}(X) \), this translates to \( m_X \circ \mathcal{S}(m_X) = m_X \circ m_S(X) \). Because the equation \( m_X \circ \varepsilon_X = id_{\mathcal{S}(X)} \) is easily established through a simple computation, the defining diagrams are commutative.

Since
\[
m_X(\alpha_1 \cdot \mu_1 + \cdots + \alpha_n \cdot \mu_n) = \alpha_1 \cdot m_X(\mu_1) + \cdots + \alpha_n \cdot m_X(\mu_n),
\]
the positive convex structure induced on \( \mathcal{S}(X) \) by this algebra is the natural one. \( \square \)

3.5.2 The Finite Case

The finite case can also easily be characterized: there are no algebras for \( \{1, \ldots, n\} \) unless \( n = 1 \). This will be shown now. Since the base space needs to be connected for entertaining an algebra, we obtain a simple geometric description as a necessary condition for the existence of algebras as a byproduct.

We need a wee bit elementary topology for this.

Definition 3.5.2 A metric space \( A \) is called connected iff the decomposition \( A = A_1 \cup A_2 \) with disjoint open sets \( A_1, A_2 \) implies \( A_1 = \emptyset \) or \( A_2 = \emptyset \).

Thus a connected space cannot be decomposed into two non-trivial open sets, so that the only clopen sets are the empty set and the space itself. The connected subspaces of the real line \( \mathbb{R} \) are just the open, half-open or closed finite or infinite intervals. The rational numbers \( \mathbb{Q} \) are not connected. A subset \( \emptyset \neq A \subseteq \mathbb{N} \) of the natural numbers which carries the discrete topology (because we assume that it is a Polish space) is connected as a subspace iff \( A = \{n\} \) for some \( n \in \mathbb{N} \).

The following facts about connected spaces are well known, cp. [32, Chapter 6.1], or any other standard reference to topology.

Lemma 3.5.3 Let \( A \) be a metric space.

1. If \( A \) is connected, and \( f : A \to B \) is a continuous and surjective map to another metric space \( B \), then \( B \) is connected.

2. If two arbitrary points in \( A \) can be joined through a connected subspace of \( A \), then \( A \) is connected.

\( \square \)

This has as a consequence a geometric description of the space underlying a monad.

Corollary 3.5.4 If \( (X, h) \) is an algebra for the Giry monad, then \( X \) is connected.
3.5 Examples

**Proof** If \( \mu_1, \mu_2 \in \mathcal{S}(X) \) are arbitrary probability measures on \( X \), then the line segment \( \{ c \cdot \mu_1 + (1 - c) \cdot \mu_2 \mid 0 \leq c \leq 1 \} \) is a connected subspace which joins \( \mu_1 \) and \( \mu_2 \). This is so because it is the image of the connected unit interval \([0, 1]\) under the continuous map \( c \mapsto c \cdot \mu_1 + (1 - c) \cdot \mu_2 \). Thus \( \mathcal{S}(X) \) is connected by Lemma 3.5.3. Since \( h \) is onto, its image \( X \) is connected. \( \dashv \)

Consequently it is hopeless to search for algebras for, say, the natural numbers or a non-trivial subset of it:

**Proposition 3.5.5** A subspace \( A \subseteq \mathbb{N} \) has an algebra for the Giry monad iff \( A \) is a singleton set.

**Proof** It is clear that a singleton set has an algebra. Conversely, if \( A \) has an algebra, then \( A \) is connected by Lemma 3.5.4, and this can only be the case when \( A \) is a singleton.

\( \dashv \)

3.5.3 The Unit Interval

The next example deals with the unit interval.

**Proposition 3.5.6** The map

\[
h : \mathcal{S}([0, 1]) \ni \mu \mapsto \int_0^1 t \mu(dt) \in [0, 1]
\]

defines an algebra \( \langle [0, 1], h \rangle \).

**Proof** In fact, \( h(\mu) \in [0, 1] \) because \( \mu \) is a sub-probability measure. It is clear that \( h(\delta_x) = x \) holds, and — by the very definition of the weak topology — that \( \mu \mapsto h(\mu) \) is continuous. Thus it remains to show by Proposition 3.2.6 that the partition induced by \( h \) is positive convex. This is a fairly simple calculation. Consequently, the partition induced by \( h \) is a G-partition, showing that \( h \) is indeed the morphism part of an algebra. It is not difficult to see that the positive convex structure induced on \([0, 1]\) is the natural one.

\( \dashv \)

This is the only algebra that has an integral representation through Lebesgue measure: suppose that

\[
h^*(\mu) = \int_0^1 f(t) \mu(dt)
\]

for some continuous \( f \). Then \( h^*(\delta_x) = f(x) \), from which \( f(x) = x \) is inferred for each \( x \in [0, 1] \).

3.5.4 Barycenter

The final example has a more geometric touch to it and deals only with the probabilistic case. We work with bounded and closed subsets of some Euclidean space and show that the construction of a barycenter yields an algebra. Fix \( X \subseteq \mathbb{R}^n \) as a bounded, closed and convex subset of the Euclidean space \( \mathbb{R}^n \) (for example, \( X \) could be a closed ball or a cube in \( \mathbb{R}^n \)).
Denote for two vectors $x, x' \in \mathbb{R}^n$ by

$$x \cdot x' := \sum_{i=1}^{n} x_i \cdot x'_i$$

their inner product. Then $\lambda x \cdot x'$ constitutes a continuous linear map on $\mathbb{R}^n$ for fixed $x'$. In fact, each linear functional on $\mathbb{R}^n$ can be represented in this way.

**Definition 3.5.7** The vector $x^* \in \mathbb{R}^n$ is called a barycenter of the probability measure $\mu \in \mathcal{P}(X)$ iff

$$x \cdot x^* = \int_X x \cdot y \mu(dy)$$

holds for each $x \in X$.

Because $X$ is compact, the integrand is bounded on $X$, thus the integral is always finite. Since the proofs for the existence and membership properties of a barycenter would lead us too far from our path of investigating stochastic relations (we would have to study the geometry of compact convex sets), we refer the reader to the literature. Basic facts about barycenters can be found e.g. in the massive overview of measure theory assembled by Fremlin [36].

**Proposition 3.5.8** The barycenter of $b(\mu)$ of $\mu \in \mathcal{P}(X)$ exists, it is uniquely determined, and it is an element of $X$. $\langle X, b \rangle$ is an algebra for the Giry monad.

**Proof** 0. Once we know that the barycenter exists, uniqueness follows from the well-known fact that the linear functionals on $\mathbb{R}^n$ separate points. The existence of the barycenter is established in [36, Theorem 461 E], its membership in $X$ follows from [36, Theorem 461 H]. Granted this, the proof that $\langle X, b \rangle$ is an algebra is sketched now along the following lines.

1. From the construction and the uniqueness of the barycenter it is clear that $b(\delta_x) = x$ holds for each $x \in X$.
2. Assume that $(\mu_n)_{n \in \mathbb{N}}$ is a sequence in $\mathcal{P}(X)$ with $\mu_n \rightharpoonup_w \mu_0$. Put $x^*_n := b(\mu_n)$ as the barycenter of $\mu_n$, then $(x^*_n)_{n \in \mathbb{N}}$ is a sequence in the compact set $X$, thus has a convergent subsequence (which we take w.l.g. as the sequence itself). Let $x^*_0$ be its limit. Then we have for all $x \in X$:

$$x \cdot x^*_n = \int_X x \cdot y \mu_n(dy) \to \int_X x \cdot y \mu_0(dy) = x \cdot x^*_0$$

Hence $b$ is continuous. Approximating $\mu$ through a convex combination of discrete measures, the above argumentation together with the convexity of $X$ shows also that $b(\mu) \in X$.

3. It remains to show that the partition induced by $b$ is convex. This, however, follows immediately from the linearity of $y \mapsto \lambda x \cdot y$. ⊣

Calculating the convex structure for $b$, we infer from affinity of the integral as a function of the measure and from

$$x \cdot b(\mu) = \int_X x \cdot y \mu(dy)$$
that \((0 \leq c \leq 1, \mu_i \in \mathcal{P}(X))\)

\[b(c \cdot \mu_1 + (1 - c) \cdot \mu_2) = c \cdot b(\mu_1) + (1 - c) \cdot b(\mu_2)\]

that the convex structure induced by \(b\) is the natural one.

It should be mentioned that this example can be generalized considerably to metrizable topological vector spaces [38, 36]. The terminological effort is, however, somewhat heavy, and the example remains essentially the same. Thus we refrain from a more general discussion.

Although the characterization of algebras in terms of positive convex structures yields a somewhat uniform approach, it becomes clear from these examples that the specific instances of the algebras provide a rather colorful picture unified only through the common abstract treatment.

### 3.6 The Left Adjoint

The identification of the algebras for the Giry monad and the observation from Proposition 3.5.1 that \(\langle \mathcal{G}(X), m_X \rangle\) is always an algebra puts us in a position where we are able to identify the left adjoint for the forgetful functor \(U : \text{Alg} \to \text{Pol}\). Define

\[\mathcal{L}(X) := \langle \mathcal{G}(X), m_X \rangle,\]

for a Polish space \(X\), and put

\[\mathcal{L}(f) := \mathcal{G}(f),\]

for the continuous map \(f : X \to Y\). Then we know from Proposition 3.5.1 that \(\mathcal{L}(X)\) is an algebra. From Lemma 3.2.1 we see that \(\mathcal{L}(f) : \mathcal{L}(X) \to \mathcal{L}(Y)\), is a morphism in \(\text{Alg}\), and since \(m : \mathcal{G}^2 \to \mathcal{G}\) is a natural transformation, \(\mathcal{L}(f)\) is an algebra morphism. Thus \(\mathcal{L} : \text{Pol} \to \text{Alg}\) is a functor.

We will write as usual \(C(a,b)\) for the morphisms \(a \to b\) in category \(C\).

### Lemma 3.6.1

Let \(\theta : \mathcal{L}(X) \to \langle Y, h \rangle\) be a morphism in \(\text{Alg}\), and put \(\Theta(\theta)(x) := \theta(\delta_x)\). This defines a bijection \(\Theta : \text{Alg}(\mathcal{L}(X), \langle Y, h \rangle) \to \text{Pol}(X, Y)\).

**Proof.**

1. Since \(x \mapsto \delta_x\) defines a continuous map \(X \to \mathcal{G}(X)\), and since the morphisms in \(\text{Alg}\) are continuous as well, \(\Theta(\theta) \in \text{Pol}(X, Y)\) whenever \(\theta \in \text{Alg}(\mathcal{L}(X), \langle Y, h \rangle)\).
2. Now suppose that \(\Theta(\theta_1(x)) = \Theta(\theta_2(x))\) holds for all \(x \in X\), thus \(\theta_1(\delta_x) = \theta_2(\delta_x)\) for all \(x \in X\). Let \(\tau = \sum_{i=1}^{m} \alpha_i \cdot \delta_{x_i}\) be a discrete sub-probability measure, then

\[
\theta_1(\tau) = \theta_1 \left( \sum_{i=1}^{m} \alpha_i \cdot \delta_{x_i} \right) \\
= \sum_{1 \leq i \leq m} \alpha_i \cdot \theta_1(\delta_{x_i}) \\
= \sum_{1 \leq i \leq m} \alpha_i \cdot \theta_2(\delta_{x_i}) \\
= \theta_2(\tau).
\]
Here $\mathcal{P}$ is the positive convex structure associated with the algebra $\langle Y, h \rangle$ by Proposition 3.4.2. Thus $\theta_1$ agrees with $\theta_2$ on all discrete measures. Since these measures are dense in the weak topology, and since $\theta_1$ as well as $\theta_2$ are continuous, we may conclude that $\theta_1(\tau) = \theta_2(\tau)$ holds for all $\tau \in \mathcal{S}(X)$. Thus $\Theta$ is injective.

3. Let $f : X \to Y$ be continuous, and put $\tilde{\theta} := h \circ \mathcal{S}(f)$, the composition being formed in $\text{Alg}$. We claim that $\tilde{\theta} \in \text{Alg}(\mathcal{L}(X), \langle Y, h \rangle)$.

In fact, consider the diagram

$$
\begin{array}{ccc}
\mathcal{S}(\mathcal{S}(X)) & \to & \mathcal{S}(Y) \\
\downarrow m_X & & \downarrow h \\
\mathcal{S}(X) & \to & Y \\
\end{array}
$$

We have

$$h \circ \mathcal{S}(\tilde{\theta}) = h \circ \mathcal{S}(h) \circ \mathcal{S}(f) = h \circ m_Y \circ \mathcal{S}(f) \quad \text{(because $\langle Y, h \rangle$ is an algebra)}
$$

$$= h \circ \mathcal{S}(f) \circ m_X \quad \text{(since $\mathcal{S}(f)$ is an $\text{Alg}$-morphism)}
$$

$$= \tilde{\theta} \circ m_X
$$

which implies that the diagram is commutative, establishing the claim. Since for each $x \in X$

$$\Theta(\tilde{\theta})(x) = h(\mathcal{S}(f)(\delta_x)) = h(\delta_{f(x)}) = f(x)
$$

we conclude that $\Theta(\tilde{\theta}) = f$, thus $\Theta$ is onto. \(\dashv\)

In order to establish the properties of an adjunction, we need to establish the naturalness of $\Theta = \Theta_{X, \langle Y, h \rangle}$, see [62, p. 80]. This means that we have to establish the commutativity of the diagrams below, given the morphisms $f \in \text{Alg}(\langle Y, h \rangle, \langle Y', h' \rangle)$ and $g \in \text{Pol}(X', X)$.

The first diagram takes care of the covariant hom-set functor $\text{Alg}(\mathcal{L}(X), \cdot) :$

$$
\begin{array}{ccc}
\text{Alg}(\mathcal{L}(X), \langle Y, h \rangle) & \overset{\Theta}{\to} & \text{Pol}(X, Y) \\
f_* & & \downarrow \Upsilon(f)_* \\
\text{Alg}(\mathcal{L}(X), \langle Y', h' \rangle) & \overset{\Theta}{\to} & \text{Pol}(X, Y) \\
\end{array}
$$

with

$$f_* : \text{Alg}(\mathcal{L}(X), \langle Y, h \rangle) \ni \theta \mapsto f \circ \theta \in \text{Alg}(\mathcal{L}(X), \langle Y', h' \rangle)$$

as composition from the left, similarly $\Upsilon(f)_*$. We see

$$\Upsilon(f)_*(\Theta(\theta))(x) = (f \circ \Theta(\theta))(x)
$$

$$= f(\Theta(\theta)(x))
$$

$$= f(\theta(\delta_x)),$$
3.7 Case Study: Derandomization

and

\[ \Theta(f_*(\theta))(x) = f_*(\theta)(x) = f(\theta(x)), \]

hence the diagram commutes. The second diagram takes care of the contravariant hom-functor \( \text{Alg}(\cdot, \langle Y, h \rangle) \):

\[
\begin{array}{ccc}
\text{Alg}(L(X), \langle Y, h \rangle) & \xrightarrow{\Theta} & \text{Pol}(X, Y) \\
\mathcal{L}(f)^* & f^* & \\
\text{Alg}(L(X'), \langle Y, h \rangle) & \xrightarrow{\Theta} & \text{Pol}(X', Y)
\end{array}
\]

Here

\[ f^* : \text{Pol}(X, Y) \ni g \mapsto g \circ f \in \text{Pol}(X', Y) \]

is composition from the right, similarly for \( \mathcal{L}(f)^* \). Because

\[ f^*(\Theta(\theta))(x') = (\Theta(\theta) \circ f)(x') = \theta(\delta_f(x')), \]

and since

\[ \Theta(\mathcal{L}(f)^*(\theta))(x') = \mathcal{L}(f)^*(\theta)(\delta_{x'}), \]

\[ = (\theta \circ \mathcal{S}(f))(\delta_{x'}), \]

\[ = \theta(\delta_{f(x')}) \]

we see that this diagram commutes as well.

Summarizing, we have established:

**Proposition 3.6.2** Define the functor \( L : \text{Pol} \to \text{Alg} \) through \( L(X) := \langle \mathcal{S}(X), m_X \rangle \) and \( L(f) := \mathcal{S}(f) \). Then \( L \) is left adjoint to the forgetful functor \( \mathcal{U} : \text{Alg} \to \text{Pol} \).

The probabilistic case is dealt with using the same arguments. The only essential place where the difference between sub-probability measures and probability measures comes formally into the discussion is in the proof of Lemma 3.6.1. Proving surjectivity of \( \Theta \), one has to take a convex combination of discrete measures, rather than a positive convex combination, as in the proof above. With this minor change all proofs carry over verbatim.

We obtain for the category \( p\text{Alg} \) of algebras for the probabilistic version of the Giry monad (the category has been introduced in Proposition 3.4.3).

**Proposition 3.6.3** Define the functor \( L_{\text{prob}} : \text{Pol} \to p\text{Alg} \) through \( L_{\text{prob}}(X) := \langle p(X), m_X \rangle \) and \( L_{\text{prob}}(f) := p(f) \). Then \( L_{\text{prob}} \) is left adjoint to the forgetful functor \( \mathcal{U} : p\text{Alg} \to \text{Pol} \).

Hence the forgetful functor on the algebras for the Giry monad has the sub-probability functor resp. the probability functor, both augmented by the monad’s multiplication, as a left adjoint. This emphasizes the close ties between positive convex resp. convex structures and probabilities and sheds further light on these functors. It also adds a formal underpinning to the intuitive understanding prevailing in Computer Science, which often expresses the probability of an outcome as a convex combination of all the possible outcomes, see e.g. [69, 92] for accounts in different fields. The interplay between convexity and probability is strikingly present for example in Heckmann’s work [43] (in fact, he often interchanges both), but surprisingly not made explicit.
3.7 Case Study: Derandomization

Derandomizations are introduced as the Eilenberg-Moore algebras for the corresponding monad, and it is argued that just the laws of an algebra provide the properties necessary for derandomization (a brief look at the Manes monad, where a similar phenomenon has been studied [4] helps in developing the argumentation).

3.7.1 Derandomizations as Algebras

In fact, let us look at modelling a nondeterministic computation using set theoretic relations. A decision $\Delta$ (borrowing the term from [4]) extracts an optimal element from a set of possible outcomes. It can be described through the following conditions, given a sup-complete partial order:

1. $\Delta(\{x\}) = x$ for all $x \in X$,
2. $\Delta(A) = \sup A$ for all $A \subseteq X$,
3. $\Delta(\bigcup D) = \sup \{\Delta(A) \mid A \in D\}$ for all $D \subseteq \mathcal{P}(X)$

Condition 1 is fairly obvious, condition 2 models selecting the optimal element w.r.t. the given partial order, and condition 3 indicates that selecting the optimal element from a collection is tantamount to optimizing choices already made for the components of the collection. Here the order is needed to make sure informally that the selection follows some consistent rule to always select the best element (whatever that means). Decisions are governed by a monad, see section 3.1.

Translating the rules above to the probabilistic case, one arrives at the following rules for selecting an element.

**Definition 3.7.1** A derandomization $h$ on the Polish space $X$ assigns each sub-probability $\tau \in \mathcal{S}(X)$ an element $h(\tau) \in X$ such that

1. $h(\delta_x) = x$, thus if the probability concentrates on a point, this is the derandomization ($\delta_x$ is the Dirac measure on $x$),
2. $h(\alpha_1 \cdot \tau_1 + \cdots + \alpha_n \cdot \tau_n) = \alpha_1 \cdot h(\tau_1) + \cdots + \alpha_n \cdot h(\tau_n)$, whenever $\tau_1, \ldots, \tau_n \in \mathcal{S}(X)$ and $\langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega$.
3. $h$ is continuous.

Condition 1 appears as fairly obvious, condition 2 implies that derandomization is geometrically smooth: if a probability is the positive convex combination of other ones, its derandomization should be, too. Continuity in condition 3 means that if one sub-probability is close to another one, their derandomizations also are. Modelling computations through a monad, we infer from Proposition 3.4.2 that derandomization carries its own positive convex structure with it, and vice versa: once a map $h$ is given, a corresponding positive convex structure can be set up, and from a positive convex structure a derandomization can be constructed. Condition 2 is replaced by a more amenable one, viz., $h \circ \mathcal{S}(h) = h \circ m_X$, where $\langle \mathcal{S}, m, e \rangle$ is the monad associated with the sub-probability
functor (see Lemma 3.2.1). In this way the operating theater is shifted to sub-probability measures on which a natural positive convex structure exists. Thus a derandomization is an algebra for this functor. Borrowing the term decision from [4] again, a derandomization may be called randomized decision, alluding to the terminology in stochastic dynamic programming, where policies select systematically from alternatives, and randomized policies do so according to some probability.

### 3.7.2 Derandomizing Ludwig

We want to interpret the characterization of the algebras in terms of the semantics for Ludwig, see section 2.4. Let \( Q \) be a positive convex structure on \( X \), hence \( Q \) yields an algebra \( h_Q \) on \( X \) according to Proposition 3.4.2. A continuous stochastic relation \( g : X \rightsquigarrow X \) is interpreted in \( Q \) through \( g^Q(x) := h_Q(g(x)) \).

The derandomization semantics \( C_Q[P] \) of a program \( P \) given a positive convex structure \( Q \) on \( X \) is \( h_Q \circ C[P] \) (with associated algebra \( h_Q \) according to Proposition 3.4.2). Following section 2.4.3, we associate with \( P \)'s work a continuous stochastic relation \( |P| \). If continuous stochastic relation \( g \) represents the initial state, the derandomization final state is given through \( C_Q[P]|g = h_Q(C[P]|g) \).

Consistency of logic and semantics translates into consistency for derandomization semantics; we see also that morphisms for strongly convex structures translate into derandomization interpretations.

#### Proposition 3.7.2

Given a program \( P \), the states \( f,g : X \rightsquigarrow X \) with \( P,f \vdash g \) and a positive convex structure \( Q \) on \( X \). Then \( C_Q[P]|g = f^Q \).

#### Proof

Immediate from consistency of the semantics, Proposition 2.4.11, and from the isomorphism between \( \text{Alg} \) and \( \text{StrConv} \), Proposition 3.4.2. \( \dashv \)

If \( \theta : Q \rightarrow R \) is a morphism between positive convex structures, define for program \( P \) the translation \( C^\theta[P] \) of the semantic function \( C[P] \) by \( \theta \) through \( C^\theta[P]|g := |P|\theta\cdot(g) \) with \( |P|\theta \cdot(g)(x)(A) := |P|(\theta^{-1}(A)) \), which is just the image of \( |P| \) under \( \mathcal{S}(\theta) \).

#### Proposition 3.7.3

Given a program \( P \), the states \( f,g : X \rightsquigarrow X \) with \( P,f \vdash g \) and positive convex structures \( Q \) and \( R \) on \( X \) with a morphism \( \theta : Q \rightarrow R \). Then

\[
\theta(C_Q[P]|g) = (C^\theta[P]|g)^R.
\]

#### Proof

Let \( h_Q \) and \( h_R \) be the algebras associated with the positive convex structures \( Q \) resp. \( R \) according to Proposition 3.4.2. Denote the composition in the base category by \( \circ \), then we have

\[
\theta(C_Q[P]|g) = \theta \circ h_Q \circ \mathfrak{m}_X \circ \mathcal{S}(P) \circ g \quad (3.10)
\]

\[
= \theta \circ h_Q \circ \mathcal{S}(h_Q) \circ \mathcal{S}(P) \circ g \quad (3.11)
\]

\[
= h_R \circ \mathcal{S}(\theta \circ h_Q) \circ \mathcal{S}(P) \circ g \quad (3.12)
\]

\[
= h_R \circ \mathcal{S}(h_R \circ \mathcal{S}(\theta)) \circ \mathcal{S}(P) \circ g \quad (3.13)
\]

\[
= h_R \circ \mathcal{S}(h_R \circ \mathcal{S}(\theta)) \circ \mathcal{S}(P) \circ g \quad (3.14)
\]

\[
= h_R \circ \mathfrak{m}_X \circ \mathcal{S}(\mathcal{S}(\theta)) \circ \mathcal{S}(P) \circ g \quad (3.15)
\]

\[
= (C^\theta[P]|g)^R \quad (3.16)
\]
The first equation (3.10) is just the definition of the Kleisli product, the second equation (3.11) makes use of the fact that \( h_Q \) defines an algebra, and the equations (3.12), (3.13), (3.14) use that \( \theta \) is a morphism for algebras. Finally, equation (3.15) refers again to the definition of the Kleisli product.

Returning to Proposition 3.7.2, establishing the counterpart to completeness is open; given the continuous derandomization \( f_1, g_1 : X \to X \) with

\[
C_Q[P]g_1 = f_1
\]

it would require finding continuous stochastic relations \( f, g : X \leadsto X \) with \( g_1 = g^Q, f_1 = f^Q \) and \( C[P]g = f \). This looks like a difficult selection problem (remember that \( f \) and \( g \) should be continuous), and conditions permitting such a selection would have to be identified.

### 3.8 Bibliographic Notes

The characterization of the algebras for the probability functor through convex structures has been known for the case that \( X \) is a compact Hausdorff space [33, 2.14] (the attribution to Swirszcz's work [89] in [33] is slightly unclear). The methods for the proof are, however, rather different: the compact case makes essential use of the right adjoint of the probability functor as a functor between the respective categories of compact Hausdorff spaces and compact convex sets. Thus Corollary 3.4.3 generalizes the known characterization to Polish spaces. In the sub-probabilistic case that is of interest here convexity is not strong enough. The positive convex structures that have been of use in other investigations could be put to use here as well; we took the definition from Pumplün's work [76].

Derandomization is a well known technique in algorithm design to gain a less randomized algorithm from a randomized one by reducing the number of random bits. This term was borrowed from this field for describing what happens when one has a probabilistic semantics and wants to see what is happening on the deterministic level. Consequently, these techniques are related in spirit.
Chapter 4

The Existence of Semi-Pullbacks

Contents

4.1 A Road Map ......................................................... 96
4.2 Extending Semi-Pullbacks of Measures ......................... 100
  4.2.1 The Compact Case ....................................... 101
  4.2.2 The General Polish Case ................................ 104
4.3 The Existence of Semi-Pullbacks ......................... 109
  4.3.1 The Polish Case ....................................... 109
  4.3.2 The Analytic Case .................................. 111
4.4 Bibliographic Notes ............................................ 112

A category is said to have semi-pullbacks if, whenever \( f : a \to b, g : c \to b \) is a pair of morphisms with the same target, there exists in the category an object \( d \) together with morphisms \( r : d \to a, s : d \to c \) such that \( f \circ r = g \circ s \). The existence of semi-pullbacks makes sure that bisimulations, which are defined as spans of morphisms, are transitive; bisimulations will be introduced in Chapter 5.

For stochastic relations, semi-pullbacks will be helpful when investigating the relationship between modal logics and bisimilar stochastic Kripke models. The argumentation goes like this: given two stochastic Kripke models \( K_1 \) and \( K_2 \) that accepts exactly the same formulas, we construct a third Kripke model \( L \) and morphisms \( \Phi_1 \) and \( \Phi_2 \) that form a co-span

\[
K_1 \xrightarrow{\Phi_1} L \xleftarrow{\Phi_2} K_2.
\]

If the category under consideration has semi-pullbacks, then the upper left corner of the diagram below may be completed

\[
\begin{array}{ccc}
M & \xrightarrow{\Psi_1} & K_1 \\
\downarrow{\Psi_2} & & \downarrow{\Phi_1} \\
K_2 & \xrightarrow{\Phi_2} & L
\end{array}
\]
From this a bisimulation is constructed. The crucial point is not the construction of $L$ (which nevertheless is quite involved in its own right) but rather the existence of the semi-pullback, i.e., the existence of the object $M$ together with the morphisms $\Psi_1$ and $\Psi_2$. This argumentation will be refined in chapter 5.

The present chapter investigates the construction of semi-pullbacks in the category of stochastic relations over Polish resp. analytic spaces. The task of construction is rather formidable and requires some non-trivial constructions, using such marvels as the Hahn-Banach Theorem and the Riesz-Representation Theorem. It capitalizes on Alexandrov’s famous embedding of a Polish space as a $G_\delta$-set (hence as a countable intersection of open sets) into a unit cube of infinite dimensions — the gist being that a Polish space can be embedded as a measurable set with compact closure into a compact set. As useful byproducts we obtain several extensions of stochastic relations from smaller to larger $\sigma$-algebras.

The problem of establishing the existence of semi-pullbacks has been solved by A. Edalat [30] for analytic spaces with universally measurable transition functions. We will refer occasionally to Edalat’s approach and comment on the differences in the Bibliographic Notes at the end of this chapter. This positive result is complemented by a consideration of extending semi-pullbacks to pullbacks or at least to weak pullbacks. It would be nice if that could be brought to work, since then various techniques that have been useful for coalgebras could be made available also for the stochastic case. Unfortunately, this cannot be the case: we conclude with the negative result that not even weak pullbacks do exist for stochastic relations. Thus we do have to invent our own techniques for exploring e.g. simple systems, rather than imitate and adapt the machinery from coalgebra.

Once the general availability of semi-pullbacks is ensured, a Hennessy-Milner Theorem on the equivalence of accepting the same formulas and bisimilarity can be established for a general class of modal logics, see chapter 6, in particular section 6.1. The investigation of bisimilarity of stochastic relations is facilitated through this result. For example it can be shown that two stochastic relations are bisimilar provided they have isomorphic factors; the converse holds under the assumption of compactness as well, so that one might speculate whether bisimilarity and having isomorphic factors are equivalent. This will be discussed in greater detail in Section 5.3.

### 4.1 A Road Map

In order to prepare for things to come, and to provide an antidote to getting the feeling that one gets lost in the measure-theoretic jungle, we will first discuss the problem and an outline of its solution in section 4.1. At the heart of the solution lies a measure extension which is provided in section 4.2, and section 4.3 constructs a solution to the existence of semi-pullbacks for stochastic relations over Polish and then over analytic spaces. This looks somewhat unrelated to the problem at hand, but it turns out that the key to solving the existence problem will be just this. The reason will become apparent soon.

Let $K = (X, Y, K)$ be a stochastic relation over the Polish spaces $X$ and $Y$. Assume that $K$ is the target of two morphisms

\[
\begin{array}{ccc}
K_1 & \xrightarrow{f_1} & K & \xleftarrow{f_2} & K_2 \\
\end{array}
\]
4.1 A Road Map

with, say, \( f_i = (\phi_i, \psi_i) \). We are looking for a stochastic relation \( L \) and two morphisms

\[
K_1 \xrightarrow{g_1} L \xrightarrow{g_2} K_2
\]

such that \( f_1 \circ g_1 = f_2 \circ g_2 \) holds.

An expansion of the first flat diagram in terms of the defining properties yields the following commutative diagram:

\[
X_1 \xrightarrow{\phi_1} X \xleftarrow{\phi_2} X_2
\]

\[
\begin{array}{ccc}
K_1 & & K_2 \\
\downarrow \ & & \downarrow \\
\mathcal{S}(Y_1) & \xrightarrow{\mathcal{S}(\psi_1)} & \mathcal{S}(Y) & \xleftarrow{\mathcal{S}(\psi_2)} & \mathcal{S}(Y_2) \\
\end{array}
\]

Written as a comprehensive diagram, the second flat diagram entails that in addition to \( f_1 \circ g_1 = f_2 \circ g_2 \) these diagrams should commute:

\[
X_1 \xrightarrow{\alpha_1} A \xleftarrow{\alpha_2} X_2
\]

\[
\begin{array}{ccc}
K_1 & & K_2 \\
\downarrow \ & & \downarrow \\
\mathcal{S}(Y_1) & \xrightarrow{\mathcal{S}(\beta_1)} & \mathcal{S}(B) & \xleftarrow{\mathcal{S}(\beta_2)} & \mathcal{S}(Y_2) \\
\end{array}
\]

Here \( L = (A, B, L) \) is the relation involved, and \( g_i = (\alpha_i, \beta_i) \) is the morphism \( g_i : L \to K_i \) for \( i = 1, 2 \). We will define

\[
A := \{ (x_1, x_2) \in X_1 \times X_2 \mid \phi_1(x_1) = \phi_2(x_2) \}
\]

\[
B := \{ (y_1, y_2) \in Y_1 \times Y_2 \mid \psi_1(y_1) = \psi_2(y_2) \},
\]

then we will argue why \( A \) and \( B \) may assumed to be Polish. Taking \( \alpha_i, \beta_i \) as the projections

\[
\begin{align*}
\alpha_i : A & \ni \langle x_1, x_2 \rangle \mapsto x_i \in X_i, \\
\beta_i : B & \ni \langle y_1, y_2 \rangle \mapsto y_i \in Y_i,
\end{align*}
\]

we need to find a stochastic relation \( L = (A, B, L) \) which makes \( g_i := (\alpha_i, \beta_i) \) into a morphism \( g_i : L \to K_i \) for \( i = 1, 2 \). Thus \( L : A \bowtie B \) should satisfy the following constraints for all \( \langle x_1, x_2 \rangle \in A \):

1. \( L(x_1, x_2) \in \mathcal{S}(B) \),
2. \( \mathcal{S}(\beta_1)(L(x_1, x_2)) = K_1(x_1) \),
3. \( \mathcal{S}(\beta_2)(L(x_1, x_2)) = K_2(x_2) \).
Reformulating again, we put
\[ \Gamma(x_1, x_2) := \{ \mu \in \mathcal{S}(B) \mid \mathcal{S}(\beta_1)(\mu) = K_1(x_1) \text{ and } \mathcal{S}(\beta_2)(\mu) = K_2(x_2) \}, \]
hence \( L(x_1, x_2) \in \Gamma(x_1, x_2) \) for all \( \langle x_1, x_2 \rangle \in A \). We want \( L : A \to \mathcal{S}(B) \) to be a measurable selector for \( \Gamma \), thus \( L : A \leadsto B \) is a stochastic relation (accounting for the measurable in measurable selector) such that \( \forall a \in A : L(a) \in \Gamma(a) \) (accounting for the selector).

Thus the problem is massaged into finding a measurable selector for the set-valued map \( \Gamma \). The existence of such a selector can be asserted under the following conditions:

1. \( \Gamma(x_1, x_2) \) is a closed subset of \( \mathcal{S}(B) \) for each \( \langle x_1, x_2 \rangle \in A \),
2. the set \( \{ \langle x_1, x_2 \rangle \in A \mid \Gamma(x_1, x_2) \cap C \neq \emptyset \} \) is a measurable subset of \( A \), whenever \( C \subseteq \mathcal{S}(B) \) is compact,
3. \( \Gamma(x_1, x_2) \neq \emptyset \) for each and any \( \langle x_1, x_2 \rangle \in A \).

These properties will characterize \( \Gamma \) as a \( C \)-measurable relation, see A.2.3.

It will not be difficult to show that property 1 is satisfied, and property 2 will also easily seen to be fulfilled. The real crucial property is the third one.

**The Crucial Point.** Fix \( \langle x_1, x_2 \rangle \in A \). We will find without much ado a measure \( \mu_0 \in \mathcal{S}(B) \) such that these equations
\[
\begin{align*}
\mathcal{S}(\beta_1)(\mu_0)(E_1) &= K_1(x_1)(E_1) \quad (4.1) \\
\mathcal{S}(\beta_2)(\mu_0)(E_2) &= K_2(x_2)(E_2) \quad (4.2)
\end{align*}
\]
are true whenever \( E_i = \psi_i^{-1}[F_i] \) for some Borel set \( F_i \subseteq Y \).

This is equivalent to saying that the measures \( \mathcal{S}(\beta_i)(\mu_0) \) and \( K_i(x_i) \) coincide on the \( \sigma \)-algebra \( \psi_i^{-1}[B(Y)] \). The latter is usually a proper sub-\( \sigma \)-algebra of \( B(Y_i) \), hence we cannot guarantee offhand equality in the equations 4.1 and 4.2 on the full Borel sets \( B(Y_i) \) with \( i = 1, 2 \). If, however, we can find a measure, say \( \mu_1 \in \mathcal{S}(B) \), such that \( \mathcal{S}(\beta_i)(\mu_1)(E_i) = K_i(x_i)(E_i) \) for all \( E_i \) ranging over all of \( B(Y_i) \), we may conclude \( \mu_1 \in \Gamma(x_1, x_2) \), and then the latter set is in fact shown to be non-empty.

Thus the problem is reduced to finding a single measure on \( B \) that has suitable marginal distributions. We will demonstrate that we can extend the marginal distributions of \( \mu_0 \) to a measure with the desired properties.

This measure extension can be done in two different ways, depending on the nature of the space \( Y \) which otherwise sits quietly in the background and serves patiently as a target space.

1. If \( Y \) is Polish, we can and do refer to a special case of Edalat’s result, and we are done. This precludes a more general solution that includes analytic spaces.

2. If \( Y \) is more generally a separable metric space, then still an extension can be found; this assumption will be crucial when the analytic case is targeted. The machinery from mathematical analysis for tackling this general case is well established, but somewhat heavy, including such the Riesz Representation Theorem and the Hahn-Banach Theorem.

It is the latter way we will propose to go here because it opens up the avenue of establishing this result also for analytic spaces.
**Discussion.** The solution provided by Edalat can be considered constructive: it works with conditional distributions and their properties. These distributions do exist essentially due to the Radon-Nikodym Theorem which in turn leans heavily on the way the Lebesgue-Daniell integral is constructed. The basic idea for the more general solution lies in a measure extension process. It works in three steps: the first is to consider the integral of the given measure, which is a linear functional, then to formulate the extension problem in terms of extending this linear functional to a linear functional on a broader domain that corresponds to our needs, and finally to represent the extended functional as an integral again. Representing the measure as an integral is easy, since only the standard techniques of integration are involved, doing the extension is technically a bit more involved and requires the Hahn-Banach Theorem, and converting the linear functional back to a measure can be done through the Riesz Representation Theorem. It is this last step that gives us the measure we are looking for.

Through the use of the Hahn-Banach Theorem we make indirectly use of Zorn’s Lemma, which in turn is known to be equivalent to the axiom of choice. Thus we propose a stronger solution but we pay more for it. This sounds probably a bit more dramatic and unusual than it really is: when constructing the product of an arbitrary family of sets or when looking into the ultrafilter extension of a Kripke model [11, Theorem 5.38] we silently take the axiom of choice for granted.

**The Road Map.** We will first delve into the problem of extending a measure with given marginal distributions; this is done in section 4.2. Then we are in a good condition for tackling the problem along the lines sketched here in section 4.3, see Figure 4.1.
4.2 Extending Semi-Pullbacks of Measures

The main argument in establishing the existence of a semi-pullback in the category of stochastic relations will be a selection argument: we will show that a certain set-valued map will have a (measurable) selector. This will require that this map always takes non-empty values. This section will be devoted to establishing a property of semi-pullbacks for measure spaces which in turn will be crucial in proving non-emptiness. Since it is rather technical in nature, it is convenient to encapsulate this development into a separate section.

We will consider the category $\text{Prob}$ of probability spaces which has as objects tuples $(X, A, \mu)$ with $\mu \in \mathcal{P}(X, A)$ for the measurable space $(X, A)$. Because for some arguments the $\sigma$-algebra is crucial, we abstain from the convention of dropping it from the notation for the space and write it down explicitly again. $\psi : (X, A, \mu) \rightarrow (Y, B, \nu)$ is a morphism in $\text{Prob}$ if $\psi : X \rightarrow Y$ is a surjective and $A - B$-measurable map which is measure preserving, i.e., $\nu = \mathcal{P}(\psi)(\mu)$ holds. $\text{Prob}$ is closed under forming products (see section A.3), in particular it contains with two object $(X, A, \mu)$ and $(Y, B, \nu)$ their product $(X \times Y, A \otimes B, \mu \otimes \nu)$, with $\mu \otimes \nu$ as the product measure which is uniquely determined through $(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B)$.

We fix for the discussion the Polish spaces $X_1$ and $X_2$ with the respective Borel sets as $\sigma$-algebras. $(Z, C)$ is assumed to be a separable measurable space. Recall that $\mathcal{F}(X, A)$ is the linear space of all $A$-$B(\mathbb{R})$-measurable bounded functions $f : X \rightarrow \mathbb{R}$; $A \mapsto \mathcal{F}(X, A)$ is monotone, hence $A \subseteq B$ implies that $\mathcal{F}(X, A)$ is a linear subspace of $\mathcal{F}(X, B)$.

Now let

$$(X_1, B(X_1), \mu_1) \xymatrix{ \psi_1 \ar[r]^\pi_1 & (Z, C, \nu) \ar[d]^{\psi_2} \ar[r]^\pi_2 \ar[d]_{\psi_1} & (X_2, B(X_2), \mu_2) \ar[d]_{\psi_2} \ar[r]^\pi_1 & (Z, C, \nu) }$$

be a pair of morphisms in $\text{Prob}$ with a common target, and assume that

\begin{align*}
(S, A, \theta) & \xymatrix{ \pi_2 \ar[r] & (X_2, B(X_2), \mu_2)} \\
\ast & \pi_1 \ar[r] & (Z, C, \nu) \ar[r] & (S, A, \theta) \ar[r] & \ast
\end{align*}

is a semi-pullback diagram in $\text{Prob}$ with

\begin{align*}
S & := \{\langle x_1, x_2 \rangle \mid \psi_1(x_1) = \psi_2(x_2)\} \in \psi_1^{-1}[C] \otimes \psi_2^{-1}[C] \\
A & := (\psi_1 \times \psi_2)^{-1}[C \otimes C].
\end{align*}

The $\pi_i$ are again the projections. The last equality addressing $A$ holds by Corollary A.2.4; thus $A$ is the smallest $\sigma$-algebra on $S$ which makes

$$\psi_1 \times \psi_2 : \langle x_1, x_2 \rangle \mapsto \langle \psi_1(x_1), \psi_2(x_2) \rangle$$

measurable.

$S$ is a Borel set, and the crucial step in the technical development will consist in “lifting” this pullback so that the object $(S, B(S), \mu)$ for some suitable $\mu \in \mathcal{P}(S, B(S))$ stands in
the upper left corner of the diagram. The essential difference is in the \(\sigma\)-algebras on \(S\): starting with the initial \(\sigma\)-algebra with respect to \(\psi_1 \times \psi_2\) we claim that we can find a measure \(\mu\) on the Borel sets of \(S\) so that the properties of a semi-pullback will be preserved.

**Proposition 4.2.1** The semi-pullback \((\star)\) in \(\text{Prob}\) may be extended to a semi-pullback

\[
(S, \mathcal{B}(S), \mu) \xrightarrow[\pi_2]{\pi_1} (X_2, \mathcal{B}(X_2), \mu_2)
\]

\[
\xrightarrow[\psi_2]{\psi_1} (X_1, \mathcal{B}(X_1), \mu_1) \rightarrow (Z, \mathcal{C}, \nu)
\]

in \(\text{Prob}\).

This entails essentially an extension process, extending \(\theta \in \mathcal{P}(S, \mathcal{A})\) to a suitable \(\mu \in \mathcal{P}(S, \mathcal{B}(S))\). We establish the existence of this extension in two steps. The first step will assume that \(X_1\) and \(X_2\) are compact Polish spaces, and the second will show how to reduce the general case to the compact one.

We will need to make precise statements regarding the measurability of a Borel map in the course of the proof. For easier reference, the technical statement below is recorded:

**Proposition 4.2.2** Let \(X\) be a Polish space, \((Y, \mathcal{B})\) be a separable measurable space, and assume that \(g : X \to Y\) is \(\mathcal{B}(X)\)-\(\mathcal{B}\)-measurable and onto. If \(f : X \to Y\) is \(\mathcal{B}(X)\)-\(\mathcal{B}\)-measurable such that \(f\) is constant on the atoms of \(g^{-1}[\mathcal{B}]\), then \(f\) is \(g^{-1}[\mathcal{B}] - \mathcal{B}\)-measurable.

**Proof** Separability implies that \(\{y\} \in \mathcal{B}\) for all \(y \in Y\). The atoms of \(g^{-1}[\mathcal{B}]\) are just the inverse images \(g^{-1}[\{y\}]\) of the points \(y \in Y\), because these sets are clearly atomic in that \(\sigma\)-algebra, and since they form a partition of \(X\). Now let \(B \in \mathcal{B}\) be a measurable set, then by assumption \(f^{-1}[B]\) is a Borel set in \(X\) which is the union of atoms of \(g^{-1}[\mathcal{B}]\). Thus the assertion follows from the Blackwell-Mackey-Theorem (Theorem A.2.6). \(\forall\)

**4.2.1 The Compact Case**

The line of attack for the case of a compact metric space will be as follows: we will construct a linear subspace of \(\mathcal{F}(S, \mathcal{B}(S))\) which contains \(\mathcal{F}(S, \mathcal{A})\) and some other functions of interest to us, and we will extend the positive linear functional \(f \mapsto \int_S f \, d\theta\) linearly to this subspace.

The next step requires the Hahn-Banach Theorem for ordered linear spaces, which may be found in [47, Lemma IX.1.4], and which is quoted here for completeness.

**Theorem 4.2.3** Assume that \((H, \leq)\) is a partially ordered vector space with a linear subspace \(H_0 \subseteq H\), and that \(L_0 : H_0 \to \mathbb{R}\) is a linear map such that \(L_0(f) \geq 0\) whenever \(0 \leq f \in H_0\). Then there exists a linear map \(L : H \to \mathbb{R}\) with the following properties:

1. \(L\) extends \(L_0\), thus if \(f \in H_0\), then \(L(f) = L_0(f)\),
2. \( L \) is positive, thus \( f \geq 0 \) implies \( L(f) \geq 0 \).

A further extension using this Hahn-Banach Theorem brings us to a positive linear functional \( \Lambda \) on \( \mathcal{F}(S, \mathcal{B}(S)) \) which then can be represented through a measure \( \mu \in \mathcal{P}(S, \mathcal{B}(S)) \), so that

\[
\Lambda(f) = \int_S f \, d\mu
\]

holds. Clearly, \( \mu \) extends \( \theta \) and is the measure we are looking for.

The commutativity of the diagram entails by the usual folklore arguments from measure theory that \((i = 1, 2)\)

\[
\forall f_i \in \mathcal{F}(X_i, \mathcal{B}(X_i)) : \int_{X_i} f_i \, d\mu_i = \int_S f_i \circ \pi_i \, d\theta
\]

holds, and by the same token it is sufficient to find an extension \( \mu \in \mathcal{P}(S, \mathcal{B}(S)) \) to \( \theta \in \mathcal{P}(S, \mathcal{A}) \) such that \((i = 1, 2)\)

\[
\forall f_i \in \mathcal{F}(X_i, \mathcal{B}(X_i)) : \int_{X_i} f_i \, d\mu_i = \int_S f_i \circ \pi_i \, d\mu
\]

holds.

**Proof** (of Proposition 4.2.1)

1. Put for \( i = 1, 2 \)

\[
\mathcal{D}_i := \{ f_i \circ \pi_i \mid f_i \in \mathcal{F}(X_i, \mathcal{B}(X_i)) \},
\]

then \( \mathcal{D}_i \subseteq \mathcal{F}(S, \mathcal{B}(S)) \), and

\[
\Lambda_0(f_i \circ \pi_i) := \int_{X_i} f_i \, d\mu_i.
\]

Then \( \Lambda_0 : \mathcal{D}_1 \cup \mathcal{D}_2 \to \mathbb{R} \) is well defined. In fact, let \( g \in \mathcal{D}_1 \cap \mathcal{D}_2 \), thus there exist functions \( f_i \in \mathcal{F}(X_i, \mathcal{B}(X_i)) \) with

\[
g = f_1 \circ \pi_1 = f_2 \circ \pi_2.
\]

We claim that \( f_1 \) is constant on the atoms of \( \psi^{-1}_1 [\mathcal{C}] \). Take \( x_1, x'_1 \in X_1 \) with \( \psi_1(x_1) = \psi_1(x'_1) \), then there exists \( x_2 \in X_2 \) such that \( \langle x_1, x_2 \rangle \in S, \langle x'_1, x_2 \rangle \in S \). Hence

\[
f_1(x_1) = g(x_1, x_2) = f_2(x_2) = g(x'_1, x_2) = f_1(x'_1).
\]

Thus \( f_1 \) is \( \psi^{-1}_1 [\mathcal{C}] \)-measurable by Proposition 4.2.2, and consequently,

\[
\int_S g \, d\theta = \int_S f_1 \circ \pi_1 \, d\theta = \int_{X_1} f_1 \, d\mu_1.
\]

Similarly,

\[
\int_S g \, d\theta = \int_{X_2} f_2 \, d\mu_2
\]

is established. This implies that \( \Lambda_0 \) is well defined.
2. Let the linear functional $\Lambda_1 : \mathcal{F}(S, A) \to \mathbb{R}$ be defined through

$$\Lambda_1(f) := \int_S f \, d\theta.$$  

We will look for a joint extension of $\Lambda_0$ and $\Lambda_1$ to the linear space spanned by $\mathcal{F}(S, A) \cup \mathcal{D}$, where $\mathcal{D} := \mathcal{D}_1 \cup \mathcal{D}_2$. This requires both functionals yielding the same value on the intersection $\mathcal{F}(S, A) \cap (\mathcal{D}_1 \cup \mathcal{D}_2)$. Assume first that $g \in \mathcal{F}(S, A) \cap \mathcal{D}_1$, thus $g = f_1 \circ \pi_1$ for some $f_1 \in \mathcal{F}(X_1, \mathcal{B}(X_1))$. Since $g$ does not depend on the second component, we may infer from the definition of $\mathcal{A}$ that $f_1$ is even $\psi_1^{-1}[\mathcal{C}]$ – measurable, hence

$$\Lambda_1(g) = \int_S g \, d\theta = \int_S f_1 \circ \pi_1 \, d\theta = \int_{X_1} f_1 \, d\mu_1 = \Lambda_0(g).$$

The argumentation for $g \in \mathcal{F}(S, A) \cap \mathcal{D}_2$ is similar.

Let $\Lambda_2$ be the joint linear extension of $\Lambda_1$ on $\mathcal{F}(S, A)$ and of $\Lambda_0$ on $\mathcal{D}$ to the linear space spanned by $\mathcal{F}(S, A)$ and $\mathcal{D}$.

From the construction it is clear that $\Lambda_2(1) = 1$ holds, and that $\Lambda_2$ is monotone.

3. The Hahn-Banach Theorem 4.2.3 for ordered linear spaces gives a positive linear operator $\Lambda : \mathcal{F}(S, \mathcal{B}(S)) \to \mathbb{R}$ that extends $\Lambda_2$. Since each continuous and bounded map $f : X_1 \times X_2 \to \mathbb{R}$ becomes a member of $\mathcal{F}(S, \mathcal{B}(S))$ when restricted to $S$, we obtain a positive linear operator $\Lambda'(f) := \Lambda(f|_S)$ on the linear space of all continuous maps $X_1 \times X_2 \to \mathbb{R}$.

Because $X_1 \times X_2$ is compact, the famous Riesz Representation Theorem yields a probability measure

$$\mu' \in \mathfrak{P}(X_1 \times X_2, \mathcal{B}(X_1 \times X_2))$$

with

$$\Lambda'(f) = \int_{X_1 \times X_2} f \, d\mu' = \int_S f \, d\mu'$$

for each $f \in \mathcal{F}(X_1 \times X_2, \mathcal{B}(X_1 \times X_2))$. Define for $B \in \mathcal{B}(S)$ the measure $\mu$ through restricting $\mu'$ to $\mathcal{B}(S)$, thus $\mu(B) := \mu'(B \cap S)$, then $\mu \in \mathfrak{P}(S, \mathcal{B}(S))$ will now be shown the measure we are looking for.

4. Let $f \in \mathcal{F}(S, A)$, then

$$\int_S f \, d\theta = \Lambda_1(f) = \Lambda_2(f) = \Lambda'(f) = \int_S f \, d\mu,$$

thus $\mu$ extends $\theta$. Let $f_i \in \mathcal{F}(X_i, \mathcal{B}(X_i))$, then $f_i \circ \pi_i \in \mathcal{D}_i \subseteq \mathcal{D}$, hence

$$\int_{X_i} f_i \, d\mu_i = \Lambda_0(f_i \circ \pi_i) = \Lambda_2(f_i \circ \pi_i) = \Lambda'(f_i \circ \pi_i) = \int_S f_i \circ \pi_i \, d\mu,$$

rendering the diagram commutative.

\[\square\]

The compactness assumption was used in the proof only to establish the existence of a measure, given a suitable linear functional on the space of continuous functions. This functional is then represented as the integral for this measure through the Riesz Theorem.
4.2.2 The General Polish Case

In the general case we do not have the Riesz Representation Theorem directly at our disposal, but compactness may nevertheless be capitalized upon since each Polish space may be embedded into a compact metric space as a measurable subspace. The famous characterization of Polish spaces due to Alexandrov [88, Remark 2.2.8] states that a topological space is Polish iff it is homeomorphic to a $G_\delta$-subset of $[0, 1]^\mathbb{N}$. In particular, a Polish space is a measurable and dense subset of a compact metric space. We will capitalize on this: $X_1$ and $X_2$ will be embedded into compact metric spaces, and this embedding will take $\psi_1, \psi_2$ and the measure $\theta$ with it. We then apply the extension procedure for the compact case. Restricting what we got from there to the original scenario, we conclude that the assertion holds also for the non-compact case.

**Proof** (of Proposition 4.2.1)

1. $X_i$ is a dense measurable subset of a compact metric space $\tilde{X}_i$ by [51, Theorem 4.14], and $\psi_i : X_i \to Z$ may be extended to a Borel measurable map $\tilde{\psi}_i : \tilde{X}_i \to Z$ by [88, Proposition 3.3.4].

Define $\tilde{\mu}_i(B_i) := \mu_i(B_i \cap X_i)$ for $B_i \in B(\tilde{X}_i)$, and put

$$S_0 := \{(x_1, x_2) \in \tilde{X}_1 \times \tilde{X}_2 \mid \tilde{\psi}_1(x_1) = \tilde{\psi}_2(x_2)\}.$$

Then $S_0 = (\tilde{\psi}_1 \times \tilde{\psi}_2)^{-1}\left[\Delta_{\tilde{X}_1 \times \tilde{X}_2}\right]$, thus

$$S_0 \in (\tilde{\psi}_1 \times \tilde{\psi}_2)^{-1}[B(Z \times Z)]$$

$$= \tilde{\psi}_1^{-1}[C] \otimes \tilde{\psi}_2^{-1}[C].$$

Since $X_i \in \tilde{\psi}_i^{-1}[C]$, and since $S = S_0 \cap (X_1 \times X_2)$, we see that $S \in \tilde{\psi}_1^{-1}[C] \otimes \tilde{\psi}_2^{-1}[C]$. Now put $\tilde{\theta}(E) := \theta(E \cap S)$ for $E \in \tilde{\psi}_1^{-1}[C] \otimes \tilde{\psi}_2^{-1}[C]$, then $\tilde{\theta}(S_0 \setminus S) = 0$, because $\tilde{\theta}$ is concentrated on $S$.

2. The construction shows that

$$\begin{array}{ccc}
(S_0, \tilde{\mu}_0, \tilde{\theta}) & \xrightarrow{\tilde{\pi}_2} & (\tilde{X}_2, B(\tilde{X}_2), \tilde{\mu}_2) \\
\downarrow \tilde{\pi}_1 & & \downarrow \tilde{\psi}_2 \\
(\tilde{X}_1, B(\tilde{X}_1), \tilde{\mu}_1) & \xrightarrow{\tilde{\psi}_1} & (Z, C, \nu)
\end{array}$$

commutes, where

$$A_0 := \left(\tilde{\psi}_1^{-1}[C] \otimes \tilde{\psi}_2^{-1}[C]\right) \cap S_0.$$

The compact case applies, hence we can find an extension $\tilde{\mu} \in \mathcal{P}(S_0, B(S_0))$ for $\tilde{\theta} \in \mathcal{P}(S_0, B(S_0))$. 

104
4.2 Extending Semi-Pullbacks of Measures

\[ \mathcal{P}(S_0, A_0) \] which lets this diagram commute:

\[
\begin{array}{ccc}
(S_0, B(S_0), \bar{\mu}) & \xrightarrow{\pi_2} & (\tilde{X}_2, B(\tilde{X}_2), \tilde{\mu}_2) \\
\tilde{\pi}_1 & \downarrow & \tilde{\psi}_2 \\
(\tilde{X}_1, B(\tilde{X}_1), \tilde{\mu}_1) & \xrightarrow{\psi_1} & (Z, \mathcal{C}, \nu)
\end{array}
\]

3. We now roll back compactification. Put for the Borel set \( B \subseteq S \)

\[
\mu(B) := \bar{\mu}(B \cap S),
\]

then \( \mu \in \mathcal{P}(S, B(S)) \), since

\[
\bar{\mu}(S_0 \setminus S) = \tilde{\theta}(S_0 \setminus S) = 0.
\]

The other properties are obvious, so that we are done with the general case, too. \( \square \)

The crucial point in this argumentation has been to prevent any mass from vanishing, i.e., to see that \( \mu(S) = 1 \) holds, which in turn could be established from the fact that \( \bar{\mu} \) extends \( \tilde{\theta} \), and for which the incorporation of \( \mathcal{F}(S, A) \) into the extension process was responsible.

We reformulate Proposition 4.2.1 in terms of sub-probability distributions. It states that there exists sometimes a common distribution for two random variables with values in a Polish space with preassigned marginal distributions. This is a cornerstone for the construction leading to the proof of Theorem 4.3.1, it shows in particular where Edalat’s work could enter the present discussion.

**Proposition 4.2.4** Let \( X_1, \) and \( X_2 \) be Polish spaces, \((Z, \mathcal{C})\) a separable measurable space, and assume that

\[
\psi_i : X_i \to Z \ (i = 1, 2)
\]

are measurable and surjective maps. Define

\[
S := \{ (x_1, x_2) \in X_1 \times X_2 \mid \psi_1(x_1) = \psi_2(x_2) \},
\]

endow \( S \) with the trace \( B(S) \) of the product \( \sigma \)-algebra, and assume that sub-probability measures \( \mu_1 \in \mathcal{G}(X_1), \mu_2 \in \mathcal{G}(X_2), \theta \in \mathcal{G}(S) \) are given such that

\[
\forall E_i \in \psi_i^{-1}[\mathcal{C}] : \mathcal{G}(\pi_i)(\theta)(E_i) = \mu_i(E_i) \ (i = 1, 2)
\]

holds, where \( \pi_1 : S \to X_1, \pi_2 : S \to X_2 \) are the projections. Then there exists \( \mu \in \mathcal{G}(S) \) such that

\[
\forall E_i \in B(X_i) : \mathcal{G}(\pi_i)(\mu)(E_i) = \mu_i(E_i) \ (i = 1, 2)
\]

holds.

**Proof** 1. We want to apply Proposition 4.2.1, so we need to show how to construct diagram (*). From the assumption we see that

\[
\begin{align*}
\mu_1(X_1) &= \theta(\pi_1^{-1}[X_1]) \\
&= \theta(S \cap (X_1 \times X_2)) \\
&= \theta(S),
\end{align*}
\]
similarly for $\mu_2$, so that $\mu_1(X_1) = \mu_2(X_2) = \theta(S)$. If $\theta(S) = 0$ the assertion is pretty obvious, so we may assume that $\theta(S) > 0$, hence it is no loss of generality to assume that all measures are probability measures.

2. Let $C \in C$, then

$$\mu_1(\psi_1^{-1}[C]) = \mathfrak{P}(\pi_1)(\theta)(\psi_1^{-1}[C]) = \mathfrak{P}(\pi_2)(\theta)(\psi_2^{-1}[C]) = \mu_2(\psi_2^{-1}[C]),$$

since $\pi_1 \circ \psi_1 = \pi_2 \circ \psi_2$ holds on $S$. So put $\nu(C) := \mu_1(\psi_1^{-1}[C])$, then $\nu \in \mathfrak{P}(Z,C)$ such that $\psi_i : (X_i,B(X_i),\mu_i) \to (Z,C,\nu)$ is a morphism in $\text{Prob}$ for $i = 1, 2$. The assumption implies that $\pi_i : (S,A,\theta) \to (X_i,B(X_i),\mu_i)$ is a morphism for $i = 1, 2$, where $A$ is the trace of the $\sigma$-algebra $\psi_1^{-1}[C] \otimes \psi_2^{-1}[C]$ on $S$. Consequently the assertion follows from Proposition 4.2.1. \[\square\]

In important special cases, there are other ways of establishing the Proposition, as will be discussed briefly.

**Remark:** 1. If $Z$ is also a Polish space, and if $\psi_i : X_i \to Z$ are bijections, then the Blackwell-Mackey Theorem (Theorem A.2.6) shows that $\psi_i^{-1}[C] = B(X_i)$. In this case the given measure $\theta \in \mathfrak{S}(S)$ is the desired one. This is so since the trace $\sigma$-algebra $A$ equals $B(S)$:

$$A = S \cap \psi_1^{-1}[C] \otimes \psi_2^{-1}[C] = S \cap B(X_1) \otimes B(X_2) = S \cap B(X_1 \times X_2) = B(S),$$

hence $\theta$ has the desired properties on the proper $\sigma$-algebra.

2. The maps $\psi_i : X_i \to Z$ are morphisms in Edalat’s category of probability measures on Polish spaces [30], provided $Z$ is a Polish space. The assertion can then be deduced from tracing the development in [30, Cor. 5.4]. The proof given above applies to Edalat’s situation as well, but it should be clear that the present proof is independent of Edalat’s. The development for the latter one depends substantially on the theory of regular conditional probabilities on analytic spaces, so that the impression might arise that the existence of the measure in question depends on these probabilities, too. The proof for Proposition 4.2.1 shows that this is not the case, that rather a straightforward proof can be given. Hence we are in the lucky position of having two independent proofs for the Polish case. Which one is preferred is largely a matter of taste: Edalat’s proof working in analytic spaces, or the one proposed here depending on the Hahn-Banach Theorem as a classical tool in analysis (but making use of the sometimes dreaded axiom of choice).

We will need an extension theorem for stochastic relations in order to secure the existence of semi-pullbacks for analytic spaces. We begin with a statement on the extension of a probability measure on a sub-$\sigma$-algebra. Note that we do not claim the uniqueness of the extension. This is different from the usual measure extensions in measure theory.
Lemma 4.2.5 Let $A$ be a sub-$\sigma$-algebra of the Borel sets of a Polish space $X$, and assume that $\theta$ is a probability measure on $A$. Then $\theta$ can be extended to a probability measure on all of $\mathcal{B}(X)$.

Proof 0. We need only to sketch the proof, since the main work has already been done in the proof of Proposition 4.2.1. Although the assertion is a bit different, the pattern of the argumentation is very similar to the one presented already.

1. First $X$ is assumed to be compact, then a combination of the Hahn-Banach-Theorem and the Riesz Representation Theorem yields the existence of the desired measure.
2. If $X$ is not compact, it is embedded as above as a measurable subset into a compact metric space. There the existence of an extension is established, and exactly the same technique as above moves that measure to the Borel sets of $X$. ⊣

The application interesting us here is the possibility to establish an extension to probabilistic relations. Before we state and prove a corresponding property, we remark that each probability on the product of two Polish spaces can be decomposed into a stochastic relation and a measure on one of the factors. We will capitalize heavily on this property when discussing the converse of a stochastic relation in section 5.6.1.

Proposition 4.2.6 Let $X$ and $Y$ be Polish spaces, and $\nu \in \mathcal{S}(X \times Y)$. Then there exists a stochastic relation $K : X \rightsquigarrow Y$ such that for all $D \in \mathcal{B}(X \times Y)$

$$\nu(D) = \int_X K(x)(D_x) \mathcal{S}(\pi_X)(\nu)(dx)$$

holds, where $\pi_X$ is the projection from $X \times Y$ to $X$.

Proof [73, Theorem V.8.1]. ⊣

The stochastic relation $K$ constructed in Proposition 4.2.6 by disintegration is uniquely determined up to sets of $\mathcal{S}(\pi_X)(\nu)$-measure zero; it is known as the regular conditional distribution of $\pi_Y$ given $\pi_X$, cf. [73, Chapter V.8].

But let us continue with the discussion of our extension problem. We will combine disintegration and the possibility of extending a measure from a sub-$\sigma$-algebra to a larger one in order to obtain

Proposition 4.2.7 Let $X$ and $Y$ be Polish spaces, assume that $\mathcal{B} \subseteq \mathcal{B}(Y)$ is a countably generated $\sigma$-algebra, and let $K_0 : (X, \mathcal{B}(X)) \rightsquigarrow (Y, \mathcal{B})$ be a stochastic relation. Then $K_0$ can be extended to a stochastic relation $K : (X, \mathcal{B}(X)) \rightsquigarrow (Y, \mathcal{B}(Y))$.

Proof 0. We will construct a probability measure on the product $\mathcal{B}(Y) \otimes \mathcal{B}$, extend this measure and then obtain the desired extension to the probabilistic relation through disintegration.

1. Let $\mu$ be a probability measure on $\mathcal{B}(X)$, and define for $D \in \mathcal{B}(Y) \otimes \mathcal{B}$ the measure

$$\mu_0(D) := \int_X K_0(x)(D_x) \mu(dx),$$

where, as usual, $D_x := \{y \in Y \mid \langle x, y \rangle \in D\}$, and by standard arguments $D_x \in \mathcal{B}$ for any $D \in \mathcal{B}(Y) \otimes \mathcal{B}$. Let $\mu_1$ be an extension of $\mu_0$ to all of $\mathcal{B}(X \times Y)$. This extension exists by
Lemma 4.2.5. Since \( \mu_1 \) is a measure on the product of two Polish spaces, there exists by Proposition 4.2.6 a stochastic relation \( K_1 : (X, \mathcal{B}(X)) \sim (Y, \mathcal{B}(Y)) \) such that

\[
\mu_1(D) = \int_X K_1(x)(D_x) \mu(dx)
\]

holds for all \( D \in \mathcal{B}(X \times Y) \). \( K_1 \) is not uniquely determined, so we have to smooth this relation somewhat.

2. Let \( \mathcal{B}_0 := \{B_n \mid n \in \mathbb{N}\} \) be a countable generator of the \( \sigma \)-algebra \( \mathcal{B} \); we may and do assume that \( \mathcal{B}_0 \) is closed under finite intersections (otherwise form all finite intersections of elements of \( \mathcal{B}_0 \), then this is still a countable generator with the desired property). Now let \( E \in \mathcal{B} \), then \( \mu_0(E \times Y) = \mu_1(E \times Y) \) holds by the construction of this extension, thus there exists for each \( E \in \mathcal{B} \) a set \( N(E) \in \mathcal{B}(X) \) with \( \mu(N(E)) = 0 \) such that

\[
\forall x \in X \setminus N(E) : K_0(x)(E) = K_1(x)(E).
\]

Now put

\[
N := \bigcup_{n \in \mathbb{N}} N(B_n)
\]

as the set of all possibly violating \( x \), then \( N \in \mathcal{B}(X) \), and \( \mu(N) = 0 \) holds.

4. We claim that for any \( x \in X \setminus N \) the equality \( K_0(x)(E) = K_1(x)(E) \) holds for every Borel set \( E \in \mathcal{B} \). In fact, put

\[
\mathcal{E} := \{E \in \mathcal{B} \mid \forall x \in X \setminus N : K_0(x)(E) = K_1(x)(E)\},
\]

then \( \mathcal{B}_0 \subseteq \mathcal{E} \) by construction, \( \mathcal{E} \) contains \( Y \), and \( \mathcal{E} \) is closed under complementation and disjoint countable unions. Thus \( \mathcal{E} = \mathcal{B} \) is inferred by the \( \pi - \lambda \)-Theorem A.1.1. Now let \( \mu_2 \) be an arbitrary probability measure on \( \mathcal{B}(Y) \), and define the stochastic relation \( K \) by cases as follows:

\[
K(x)(D) := \begin{cases} 
K_1(x)(D), & x \not\in N, D \in \mathcal{B}(Y) \\
K_0(x)(D), & x \in N, D \in \mathcal{B} \\
\mu_2(D), & x \in N, D \not\in \mathcal{B}.
\end{cases}
\]

This relation has the desired properties. \( \dashv \)

The result applies directly to making a pair of surjective and measurable maps into morphisms under a rather weak condition of separability.

**Lemma 4.2.8** Let \( M := (A, B, M) \) be a stochastic relation between the measurable spaces \( A \) and \( B \), and assume that \( B \) is separable. If \( X \) and \( Y \) are Polish spaces with measurable and surjective maps \( \phi : X \to A, \psi : Y \to B \), then there exists a stochastic relation \( K := (X, Y, K) \) which makes \( f := (\phi, \psi) \) a morphism \( f : K \to M \).

**Proof** Let \( \mathcal{B} \) the \( \sigma \)-algebra on \( B \), then \( \psi^{-1}(\mathcal{B}) \) is a countably generated sub-\( \sigma \)-algebra of \( \mathcal{B}(Y) \). Define for \( x \in X \) and \( D \in \mathcal{B} \)

\[
K_0(x)(\psi^{-1}[D]) := M(\phi(x))(D),
\]

then \( K_0 : (X, \mathcal{B}(X)) \simeq (Y, \psi^{-1}[\mathcal{B}]) \) is a stochastic relation which can be extended to a stochastic relation \( K : (X, \mathcal{B}(X)) \simeq (Y, \mathcal{B}(Y)) \) by Proposition 4.2.7. It is plain from the construction that \( \mathcal{S}(\psi) \circ K = M \circ \phi \) holds. \( \dashv \)
4.3 The Existence of Semi-Pullbacks

We will show now that semi-pullbacks exist in a rather general setting, generalizing the constructions in [30, 25]. This will ultimately lead to showing that semi-pullbacks exist for analytic objects, and it will turn out that the object underlying such a semi-pullback is Polish.

4.3.1 The Polish Case

The central Lemma reads as follows:

**Lemma 4.3.1** Let $K_i$ be Polish objects, and assume that $K = (X,Y,K)$ is a stochastic relation, where $X,Y$ are separable measurable spaces. In Stoch each diagram

\[
\begin{array}{ccc}
K_1 & \xrightarrow{f_1} & K \\
\downarrow & & \downarrow \\
K_2 & \xrightarrow{f_2} & K
\end{array}
\]

has a semi-pullback

\[
\begin{array}{ccc}
M & \xrightarrow{g_1} & K_1 \\
\downarrow & & \downarrow \\
K_2 & \xrightarrow{f_2} & K
\end{array}
\]

with a Polish object $M$.

**Proof 1.** Assume $K_i = (X_i,Y_i,K_i)$ with $f_i = (\phi_i,\psi_i), i = 1,2$. In view of Lemma A.2.3 we may and do assume that the respective $\sigma$-algebras on $X$ and $Y$ are the Borel sets of second countable metric spaces. Because of Proposition A.2.1 we may assume that the respective $\sigma$-algebras on $X_1$ and $X_2$ are obtained from Polish topologies which render $\phi_1$ and $K_1$ as well as $\phi_2$ and $K_2$ continuous. These topologies are fixed for the proof. Put

\[
A := \{ (x_1,x_2) \in X_1 \times X_2 \mid \phi_1(x_1) = \phi_2(x_2) \},
\]

\[
B := \{ (y_1,y_2) \in Y_1 \times Y_2 \mid \psi_1(y_1) = \psi_2(y_2) \},
\]

then both $A$ and $B$ are closed, hence Polish. $\alpha_i : A \to X_i$ and $\beta_i : B \to Y_i$ are the projections, $i = 1,2$. The diagrams
are commutative by assumption, thus we know that for \( x_i \in X_i \)

\[
K(\phi_1(x_1)) = \mathcal{G}(\psi_1)(K_1(x_1)),
K(\phi_2(x_2)) = \mathcal{G}(\psi_2)(K_2(x_2))
\]

both hold. The construction implies that \((\psi_1 \circ \beta_1)(y_1, y_2) = (\psi_2 \circ \beta_2)(y_1, y_2)\) is true for \(\langle y_1, y_2 \rangle \in B\), and \(\psi_1 \circ \beta_1 : B \rightarrow Y\) is surjective.

2. Fix \(\langle x_1, x_2 \rangle \in A\). Separability of the target spaces now enters: Corollary A.2.4 shows that \(\mathcal{G}(\psi \circ \beta_1)(\mu_0) = K(\phi_1(x_1))\), consequently, \(\mathcal{G}(\psi \circ \beta_1)(\mu_0) = \mathcal{G}(\psi_1)(K_1(x_i))\) \((i = 1, 2)\). But this means

\[
\forall E_i \in \psi_i^{-1}[\mathcal{B}(Y)] : \mathcal{G}(\beta_1)(\mu_0)(E_i) = K_i(x_i)(E_i) \ (i = 1, 2).
\]

Put

\[
\Gamma(x_1, x_2) := \{ \mu \in \mathcal{G}(B) \mid \mathcal{G}(\beta_1)(\mu) = K_1(x_1) \land \mathcal{G}(\beta_2)(\mu) = K_2(x_2) \},
\]

then Proposition 4.2.4 shows that \(\Gamma(x_1, x_2) \neq \emptyset\).

3. Since \(K_1\) and \(K_2\) are continuous, \(\Gamma : A \rightarrow \mathcal{P}(\mathcal{G}(B))\) is easily established. The set

\[
\exists \Gamma(C) = \{ (x_1, x_2) \in A \mid (x_1, x_2) \cap C \neq \emptyset \}
\]

is closed in \(A\) for compact \(C \subseteq \mathcal{G}(B)\). In fact, let \(\langle x_1^{(n)}, x_2^{(n)} \rangle \rangle_n \in \mathbb{N}\) be a sequence in this set with \(x_i^{(n)} \rightarrow x_i\), as \(n \rightarrow \infty\) for \(i = 1, 2\), thus \(\langle x_1, x_2 \rangle \in A\). There exists \(\mu_n \in C\) such that \(\mathcal{G}(\beta_1)(\mu_n) = K_i(x_i^{(n)})\). Because \(C\) is compact, there exists a converging subsequence \(\mu_{s(n)}\) and \(\mu \in C\) with \(\mu = \lim_{n \rightarrow \infty} \mu_{s(n)}\) in the topology of weak convergence. Continuity of \(K_i\) implies that \(\mathcal{G}(\beta_i)(\mu) = K_i(x_i)\), consequently \(\langle x_1, x_2 \rangle \in \exists \Gamma(C)\), thus this set is closed, hence measurable.

4. From Proposition A.2.7 it is now inferred that there exists a measurable map \(M : A \rightarrow \mathcal{G}(B)\) such that \(M(x_1, x_2) \in \Gamma(x_1, x_2)\) holds for every \(\langle x_1, x_2 \rangle \in A\). Thus \(M : A \rightsquigarrow B\) is a stochastic relation with

\[
K_1 \circ \alpha_1 = \mathcal{G}(\beta_1) \circ M,
K_2 \circ \alpha_2 = \mathcal{G}(\beta_2) \circ M.
\]

Hence \(M := (A, B, M)\) is the desired semi-pullback. \(\dagger\)

This statement includes several interesting special cases:

**Theorem 4.3.2** Semi-pullbacks exist in the category \(\text{PolStoch}\) of stochastic relations over Polish spaces.

**Proof** This follows immediately from Lemma 4.3.1. \(\dagger\)

**Corollary 4.3.3** Suppose that in the diagram of Lemma 4.3.1 the target object \(K\) is an analytic object. Then a Polish semi-pullback of that diagram exists in \(\text{Stoch}\).

**Proof** This also follows immediately from Lemma 4.3.1. \(\dagger\)
4.3 The Existence of Semi-Pullbacks

4.3.2 The Analytic Case

We obtain from Theorem 4.3.2 together with Corollary 4.3.3 the interesting generalization to analytic spaces through essentially an extension argument. Suppose that we have an analytic object \( M \), then we can find a Polish object \( K \) and a morphism \( g : K \to M \) by Lemma 4.2.8. This is so since analytic spaces are surjective images of Polish spaces under measurable maps, and since analytic spaces are — as measurable spaces — separable. This observation has as a somewhat unexpected consequence that semi-pullbacks do exist for analytic spaces:

**Corollary 4.3.4** Let \( M_i \) be an analytic object for \( i = 1, 2 \) and assume that \( M := (A, B, M) \) is a stochastic relation for the separable measurable spaces \( A, B \). For the pair

\[
\begin{array}{ccc}
M_1 & f_1 & \to & M & f_2 \leftarrow & M_2
\end{array}
\]

there exist both a Polish object \( K \) and morphisms

\[
\begin{array}{ccc}
M_1 & g_1 & \to & K & g_2 \leftarrow & M_2
\end{array}
\]

forming a semi-pullback.

**Proof** We can find Polish objects \( K_i \) and morphisms in Stoch extending the diagram

\[
\begin{array}{ccc}
K_1 & h_1 & \to & K_2
\end{array}
\]

\[
\begin{array}{ccc}
M_1 & f_1 & \to & M & f_2 \leftarrow & M_2
\end{array}
\]

Now, using Lemma 4.3.1, find a semi-pullback

\[
\begin{array}{ccc}
K_1 & t_1 & \to & K & t_2 \leftarrow & K_2
\end{array}
\]

for the diagram

\[
\begin{array}{ccc}
K_1 & h_1 \circ f_1 & \to & M & h_2 \circ f_2 \leftarrow & K_2
\end{array}
\]

Here \( K \) is a Polish object, and \( t_1, t_2 \) are morphisms in Stoch. Putting \( g_i := h_i \circ t_i \) for \( i = 1, 2 \) now establishes the claim. \( \Box \)

Thus we have in particular established:

**Theorem 4.3.5** The category \( \text{anStoch} \) of stochastic relations over analytic spaces has semi-pullbacks. They may be chosen as Polish objects. \( \Box \)

We close with a negative result, indicating that it is not possible to strengthen the results obtained here towards the existence of weak pullbacks or to pullbacks. Recall that a semi-pullback \( r : d \to a, s : d \to c \) for a pair of morphisms \( f : a \to b, g : c \to b \) is a weak pullback iff the following holds: whenever \( r' : d' \to a, s' : d' \to c \) forms a commutative diagram with \( f \) and \( g \) (i.e., \( f \circ r' = g \circ s' \) holds), then there exists a morphism \( h : d' \to d \) with \( r' = r \circ h, s' = s \circ h \). If morphism \( h \) is uniquely determined, then \( d \) together with \( r \) and \( s \) is called a pullback.

The category \( \text{PolProb} \) is the full subcategory of \( \text{Prob} \) (cf. section 4.2) which has probability spaces based on Polish spaces as objects.
**Proposition 4.3.6** Let \((X, \mu)\) and \((Y, \nu)\) be objects in \(\text{PolProb}\), and assume that \(\phi : (X, \mu) \to (Y, \nu)\) is a morphism in \(\text{PolProb}\) such that \(\phi : X \to Y\) is not bijective. Then the kernel pair

\[
\begin{array}{ccc}
(X, \mu) & \xto{\phi} & (Y, \nu) \\
\xleftarrow{\phi} & & \xleftarrow{\phi} \end{array}
\]

does not have a weak pullback in \(\text{PolProb}\).

**Proof** Assume that \((P, \rho)\) is a weak pullback for that kernel pair with morphisms based on the maps \(\pi_1 : P \to X\) and \(\pi_2 : P \to X\). Because the category of Polish spaces with surjective Borel maps has finite products and equalizers, we conclude that

\[
P = \{ (x_1, x_2) \in X \times X \mid \phi(x_1) = \phi(x_2) \}
\]

and that \(\pi_1, \pi_2\) are just the projections. Because the identity \(id_X : (X, \mu) \to (X, \mu)\) is plainly a morphism with \(f \circ id_X = f \circ id_X\), we find a morphism \(\chi : (X, \mu) \to (P, \rho)\) such that \(\pi_1 \circ \chi = \pi_2 \circ \chi\). Thus each element of \(P\) must be of the form \((x, x)\), contradicting the fact that \(\phi\) is not injective. \(\dashv\)

Because \(\text{PolProb}\) is a full subcategory of \(\text{PolStoch}\), we may conclude

**Corollary 4.3.7** Neither \(\text{PolStoch}\) nor \(\text{anStoch}\) have weak pullbacks. \(\dashv\)

When comparing the situation visible for the probability functor against the scenario in general coalgebras, the reader may wish to consult the survey [77] and to observe that many of the main theorems and constructions require the existence of a weak pullback for the functor governing the coalgebra. Corollary 4.3.7 is in part responsible for the fact that many constructions that are quite similar to the ones performed for coalgebras will have to develop their own, specific proof for stochastic relations, which does not agree or even resemble a coalgebraic argument. The discussion of simple systems in section 5.5 will demonstrate this rather clearly. This is also the deeper reason why modelling software architectures in categories like \(\text{PolStoch}\) will be confined to linear models, because modelling decisions in them is difficult, see [57], and the discussion in section 2.3.7.

### 4.4 Bibliographic Notes

The problem to secure the existence of a semi-pullback has been addressed in a variety of ways, varying the base category suitably:

1. Edalat [30] considers a category of Markov processes with the state space an analytic space, and the transition probability function as universally measurable. The solution to this instance of the problem is essentially based on an explicit construction using regular conditional probabilities which are available due to the Polish descent of analytic spaces.

   Once this problem is solved, it becomes possible to tackle the equivalence of bisimilarity and accepting the same formulas for a whole family of modal logics with a countable number of diamonds, all interpreted over analytic spaces with a labelled Markov transition system based on universally measurable transition functions, see [20].

---

112
2. In [25] Markov processes over Polish spaces are considered, where the transition probability function is Borel measurable. The problem is transformed into finding a measurable selector for a set-valued map, as in the present discussion.

Based on this solution, it is shown that bisimilarity and accepting the same set of formulas are equivalent for a simple negation free modal logic with a countable number of diamonds. The labelled Markov transition processes come from Borel functions based on Polish spaces, and the technical condition was that one of the processes is small, i.e., has a Borel section [88].

We did refine here the technique employed in [25] in order to show that semi-pullbacks exist in categories of Markov processes over analytic spaces when the transition probability functions are Borel measurable (rather than universally measurable). For practical purposes, the difference between universal and Borel measurability is probably negligible, for structural purposes it is not. Borel measurability is defined in terms of the inverse image of Borel sets (in exactly the same way as continuity is defined in terms of the inverse image of the open sets, or as uniform continuity is in terms of the inverse image of neighborhoods), universal measurability requires additionally the concept of completing the Borel sets through all finite measures, thus requires the additional concept of a finite measure. Hence Borel measurability is conceptually simpler and appears as more fundamental.
This chapter will introduce smooth equivalence relations as one of the central tools for investigating stochastic relations. We have seen already that smooth relations occur naturally when looking at a classification of relations through algebras for the associated monad in Section 3.1. There we had a look at an equivalence relation that occurred as the kernel of a Borel map.

We will now investigate these kernels more systematic and from different angles. As it turns out, these relations occur quite naturally in another disguise: take for example
a Kripke model $\mathcal{K}$ with state space $S$ for the simple modal logic introduced in the Introduction with its set $\Phi$ of formulas. If the set $A$ of actions and the set $AP$ of atomic propositions are countable, $\Phi$ is countable as well. Now define

\[ s \equiv s' \iff \forall \phi \in \Phi : [\mathcal{K}, s \models \varphi \iff \mathcal{K}, s' \models \varphi]. \]

Then the relation $\equiv$ is smooth, as we will show, and the invariant sets for this relation become of interest: A set $B \subseteq S$ is $\equiv$-invariant iff $s \in B$ and if $s'$ satisfies exactly the same formulas as $s$ together imply $s' \in B$, or, using $\equiv$, if $s \in B$ and $s' \equiv s$ imply $s' \in B$, thus iff $B$ is the union of the $\equiv$-equivalence classes. The invariant Borel sets form a $\sigma$-algebra which uniquely identifies the relation. A variant of this theme will be interesting as well: Say that $s \equiv_F s'$ iff $s$ and $s'$ satisfy exactly the same formulas from $F \subseteq \Phi$. When investigating a continuous time stochastic logic in 6.3, we will expand the reach of $\equiv_F$ in the sense that we are looking for a larger set $G$ of formulas so that $\equiv_F$ equals $\equiv_G$ (which means that it is enough to check the formulas in $F$ in order to collect sufficient information regarding the formulas in $G$). Although it is far from evident, the invariant sets help here, too.

We will investigate these relations in this chapter, introduce congruences for stochastic relations based on smooth relations, and show how to factor stochastic relations with congruences. Factor systems will be an important tool for the investigation of bisimulations, which will also be introduced here. First, bisimulations will be introduced as spans of morphisms with an additional property that ties the bisimilar objects together.

We will develop a sufficient condition for two stochastic relations to be bisimilar. This condition is based on congruences that in some sense are generating each other, later we will see that a natural condition like the one in the Hennessy-Milner Theorem for modal logics can be derived from it. In a special case the condition can be shown to be sufficient: this requires the mediating system to be a compact metric space. In particular the criterion applies to finite relations. It says that we need only to look at the subsystems of two stochastic relations. If we detect subsystems that are isomorphic, then we know that the systems are bisimilar.

One of the techniques for the investigation of bisimilar systems is factoring a system. This is investigated here as well, building the bridge to classical algebraic systems like groups, modules and the like. Because we have congruences at our disposal, we may look at the factor system and see which properties it has, and how the morphisms relate to factoring. One of the results will be that for a morphism $f : K \rightarrow K'$ an isomorphism between $K/\ker (f)$ and $K'$ can be established (morphisms are epis). If $d$ is a congruence on the factor relation $K/c$, then we show that $(K/c)/d$ is isomorphic to $K/c \cdot d$, where $c \cdot d$ is a congruence of $K$ that is coarser than $c$ and encodes the properties of $d$ on the level of the base system. This basically entails that all the properties of a factor system may be derived from the base system itself, so that iterative factoring does not lead to unpleasant surprises. This isomorphism is similar to the ones captured in classical group theory through the Second Isomorphism Theorem.

We deal in this chapter also with systems that are simple in the sense that they do not have any interesting subsystems. So a stochastic relation $K$ is simple if a factor system $K/c$ is either trivial or equal to $K$ itself. These systems are of considerable coalgebraic interest, because — via final systems — they form the basis for the proof principle of coinduction. If we could identify final systems in a sensible way, then we could establish
probabilistic coinduction as well. But this hope cannot be supported: final systems are much too simple to be of any interest. We identify simple systems through their bisimulations (this is truly in the coalgebraic spirit), but then have a look at what this entails for final systems, provided they exist, and here coalgebraic hope will leave us.

## 5.1 Smooth Equivalence Relations

An equivalence relation $\rho$ on an analytic space is smooth (or countably generated) iff it can be decided whether or not two elements are equivalent by looking at a countable family of Borel sets. Smooth relations have already been introduced in section 3.2.3 as the kernels of Borel measurable maps. We give a definition in terms of a determining sequence of Borel sets and relate the characterizations in Lemma 5.1.6.

**Definition 5.1.1** Let $X$ be an analytic space and $\rho$ an equivalence relation on $X$. Then $\rho$ is called smooth iff there exists a sequence $(A_n)_{n \in \mathbb{N}}$ of Borel sets such that

$$x \rho x' \iff \forall n \in \mathbb{N} : [x \in A_n \iff x' \in A_n].$$

$(A_n)_{n \in \mathbb{N}}$ is said to determine the relation $\rho$.

We obtain immediately from the definition that a smooth equivalence relation seen as a subset of the Cartesian product is a Borel set:

**Corollary 5.1.2** Let $\rho$ be a smooth equivalence relation on the analytic space $X$, then $\rho$ is a Borel subset of $X \times X$.

**Proof** Suppose that $(A_n)_{n \in \mathbb{N}}$ determines $\rho$. Since $x \rho x'$ is false iff there exists $n \in \mathbb{N}$ with $\langle x, x' \rangle \in A_n \times X \setminus A_n \cup X \setminus A_n \times A_n$, we obtain

$$(X \times X) \setminus \rho = \bigcup_{n \in \mathbb{N}} (A_n \times (X \setminus A_n)) \cup ((X \setminus A_n) \times A_n).$$

This is clearly a Borel set in $X \times X$. ⊣

The invariant Borel sets will be a powerful tool for investigating smooth relations. In order to appreciate this $\sigma$-algebra fully, we show the following [88, Lemma 3.1.6].

**Lemma 5.1.3** Assume that for a set $M$ an equivalence relation $\sim$ is defined through

$$m \sim m' \iff \forall G \in \mathcal{G} : [m \in G \iff m' \in G],$$

where the elements of $\mathcal{G}$ are subsets of $M$. Then

$$m \sim m' \iff \forall G \in \sigma(\mathcal{G}) : [m \in G \iff m' \in G].$$

**Proof** It is not difficult to see that the set

$$\mathcal{H} := \{G \in \sigma(\mathcal{G}) \mid m \in G \iff m' \in G\}$$

is a $\sigma$-algebra, where $m, m' \in M$ are fixed. For example, if $G_1, G_2 \in \mathcal{H}$ and $m \in G_1 \cup G_2$, then $m \in G_1$ or $m \in G_2$. Depending on which case applies, $m' \in G_1$ or $m' \in G_2$, thus
m' ∈ G₁ ∪ G₂, and vice versa. But by assumption G ⊆ H, thus σ(G) ⊆ H, and the conclusion follows. ⊣

This Lemma tells us that invariance with respect to an equivalence relation, which will be defined below, is carried over from a generator to its σ-algebra. This implies that we have some degrees of freedom when selecting a generator.

**Definition 5.1.4** Let ρ be a smooth equivalence relation on an analytic space X.

1. A subset A ⊆ X is called ρ-invariant iff x ∈ A and x ρ x' together imply x' ∈ A.

2. Denote by INV (B(X), ρ) the σ-algebra of ρ-invariant Borel subsets of X.

Thus a ρ-invariant set A can be written as the union of the equivalence classes of its sets, A = ∪{[x]ρ | x ∈ A}. We see further that the invariant Borel subsets constitute the σ-algebra of (Aₙ)ₙ∈ℕ which determines ρ. This will be investigated further in a moment. These are quite simple examples:

**Lemma 5.1.5** The identity relation Δₓ and the universal relation Uₓ are for each Polish space X smooth equivalence relations with

\[
INV (B(X), Δₓ) = B(X),
INV (B(X), Uₓ) = \{∅, X\}.
\]

**Proof** The assertion is trivial for the universal relation. One argues for the identity relation as follows: the Borel sets of X are countably generated, and one can find such a countable generator G that separates points. This implies that Δₓ has G as the determining family, and since σ(G) = B(X), the assertion follows. ⊣

The following characterization of smooth equivalence relations (cp. [88, Exercise 5.1.10]) is sometimes helpful and shows that it is not necessary to look only at sequences of sets. It indicates that Borel measurable maps and smooth relations are very closely connected.

**Lemma 5.1.6** Let ρ be an equivalence relation on an analytic set X. Then these conditions are equivalent:

1. ρ is smooth.

2. There exists a sequence (fₙ)ₙ∈ℕ of Borel maps fₙ : X → Z into an analytic space Z such that ρ = ∩ₙ∈ℕ ker (fₙ).

3. There exists a Borel map f : X → Y into an analytic space Y with ρ = ker (f).

**Proof** 1. ⇒ 2: Let (Aₙ)ₙ∈ℕ determine ρ, then

\[
x ρ x' ⇔ ∀n ∈ ℕ : [x ∈ Aₙ ⇔ x' ∈ Aₙ]
⇔ ∀n ∈ ℕ : χₐₙ (x) = χₐₙ (x').
\]

Thus take Y = {0, 1} and fₙ := χₐₙ.
2. $\Rightarrow$ 3: Put $Z := Y^N$. This is an analytic space in the product $\sigma$-algebra, and

$$f := \begin{cases} X \to Z \\ x \mapsto (f_n(x))_{n \in \mathbb{N}} \end{cases}$$

is Borel measurable with

$$f(x) = f(x') \iff \forall n \in \mathbb{N} : f_n(x) = f_n(x').$$

3. $\Rightarrow$ 1: Since $Z$ is analytic, it is separable, hence the Borel sets are generated through a sequence $(B_n)_{n \in \mathbb{N}}$ which separates points. Put $A_n := f^{-1}[B_n]$, then $(A_n)_{n \in \mathbb{N}}$ is a sequence of Borel sets, because the base sets $B_n$ are Borel in $Z$, and because $f$ is Borel measurable. We claim that $(A_n)_{n \in \mathbb{N}}$ determines $\rho$:

$$f(x) = f(x') \iff \forall n \in \mathbb{N} : [f(x) \in B_n \iff f(x') \in B_n] \quad (\text{since } (B_n)_{n \in \mathbb{N}} \text{ separates points in } Z)$$

$$\iff \forall n \in \mathbb{N} : [x \in A_n \iff x' \in A_n].$$

Thus each smooth equivalence relation may be represented as the kernel of a Borel map, and vice versa.

The interest in analytic spaces comes from the fact that factoring an analytic space through a smooth equivalence relation will result in an analytic space again. This requires first and foremost the definition of a measurable structure induced by the relation. The natural choice is the structure imposed by the factor map. The final $\sigma$-algebra on $X/\rho$ with respect to the Borel sets on $X$ and the natural projection $\eta_\rho$ will be chosen; it is denoted by $\mathcal{B}(X)/\rho$. Thus $\mathcal{B}(X)/\rho$ is the largest $\sigma$-algebra $\mathcal{C}$ on $X/\rho$ rendering $\eta_\rho$ a $\mathcal{B}(X)$-$\mathcal{C}$-measurable map. Then it turns out that $\mathcal{B}(X/\rho)$ coincides with $\mathcal{B}(X)/\rho$:

**Proposition 5.1.7** If $\rho$ is a smooth equivalence relation on the analytic space $X$, then $X/\rho$ is an analytic space. $\dashv$

This statement is established e.g. in [5, Chapter 3.3]. Since its proof requires some preparations that would lead the present discussion a bit far from its destination, the reader is referred to [5]. The reader who prefers to solve this problem as an exercise is referred to [88, Exercise 5.1.14] instead.

The property above is fairly fundamental for the development of the algebraic theory of stochastic relations; it is one of the reasons for sometimes preferring analytic spaces over Polish ones, since the latter ones are not closed under factoring through a smooth relation. It will enable factoring through a congruence (see section 5.2) without having to be afraid that the arising structure will loose essential properties. This will be discussed in due course.

### 5.1.1 Invariant Borel Sets

The invariant Borel sets may be characterized through the factor map by taking the inverse image of the Borel sets of a factor space. This will give a fairly practical handle on the invariant sets. The next Lemma is a bit more general by considering general surjective Borel maps, but we will see that this is helpful indeed.
Lemma 5.1.8 Let \( X, Y \) be analytic spaces, and assume that \( f : X \to Y \) is a surjective and Borel measurable map. Then \( f^{-1}[B(Y)] = \mathcal{INV}(B(X), \ker(f)) \).

**Proof** 1. Given \( A \in \mathcal{INV}(B(X), \ker(f)) \), we show first that
\[
f^{-1}[f[A]] = A
\]
holds. In fact, \( A \subseteq f^{-1}[f[A]] \) is always true. Let \( s \in f^{-1}[f[A]] \), thus \( f(s) = f(s') \) for some \( s' \in A \). Since \( A \) is \( \ker(f) \)-invariant, this implies \( s \in A \), accounting for the other inclusion.

2. Let again \( A \in \mathcal{INV}(B(X), \ker(f)) \), then \( f[A] \subseteq Y \) is analytic. We claim that
\[
f[X \setminus A] = Y \setminus f[A]
\]
holds. For, if \( y \in f[S \setminus A] \), we can find \( x \notin B \) with \( f(x) = y \). Assuming that \( y = f(x') \) for some \( x' \in A \), we would infer that \( x \in A \) due to the \( \ker(f) \)-invariance of \( A \), and since \( \langle x, x' \rangle \in \ker(f) \). This is a contradiction. This settles the non-trivial inclusion. From the representation just established we see that \( Y \setminus f[A] \) is analytic, and from Souslin’s Theorem (Theorem A.2.5) we infer now that \( f[A] \) is Borel in \( Y \).

3. It is clear that for each \( B \in B(Y) \) its inverse image \( f^{-1}[B] \) under \( f \) is a Borel set which is \( \ker(f) \)-invariant. On the other hand, if \( A \in \mathcal{INV}(B(X), \ker(f)) \), we write \( A = f^{-1}[f[A]] \) by part 1, and \( f[A] \in B(Y) \) by part 2. This implies the desired equality. \( \Box \)

As a by-product we obtain a characterization of \( \rho \)-invariant Borel sets in analytic spaces through the generating sequence \((A_n)_{n \in \mathbb{N}}\); this result is well-known for Polish spaces, cp. [88, Lemma 5.1.16], it seems to be new for the analytic case. As a consequence, we can characterize the \( \rho \)-invariant Borel set through the canoncic projection \( \eta_{\rho} \).

Proposition 5.1.9 Let \( X \) be an analytic space with a smooth equivalence relation \( \rho \), then the \( \rho \)-invariant Borel sets of \( X \) are exactly the inverse images of the canoncic projection \( \eta_{\rho} \), viz.,
\[
\mathcal{INV}(B(X), \rho) = \eta_{\rho}^{-1}[B(X/\rho)]
\]
holds. Moreover, if \( \rho \) is determined by the sequence \((A_n)_{n \in \mathbb{N}}\) of Borel sets \( A_n \subseteq X \), then
\[
\mathcal{INV}(B(X), \rho) = \sigma(\{A_n \mid n \in \mathbb{N}\}).
\]

**Proof** 1. \( X/\rho \) is an analytic space, and \( \eta_{\rho} : X \to X/\rho \) is surjective and onto. Thus the first assertion follows from Lemma 5.1.8 upon observing that \( \rho = \ker(\eta_{\rho}) \) holds.

2. Let \( A \in B(X/\rho) \) be a Borel set in \( X/\rho \). Plainly,
\[
A = \bigcup \{\{[x]_{\rho} \mid [x]_{\rho} \in A\},
\]
so it is enough to show that each \( \{[x]_{\rho}\} \) constitutes an atom in \( \sigma(\{\eta_{\rho}[A_n] \mid n \in \mathbb{N}\}) \).

3. It is easy to see that
\[
\cap\{\eta_{\rho}[A_n] \mid t \in A_n\} \cap \{X/\rho \setminus \eta_{\rho}[A_n] \mid t \notin A_n\}
\]
contains the class \([x]_\rho\) as its only element, and that \((X/\rho) \setminus \eta_\rho [A_n] = \eta_\rho [T \setminus A_n]\), because \(A_n\) is \(\rho\)-invariant, cp. part 2 of the proof of Lemma 5.1.8. Thus the atom \([(x)_\rho]\) is a member of \(\sigma ([\eta_\rho [A_n] | n \in \mathbb{N}])\).

We obtain as a Corollary that a smooth equivalence relation is determined uniquely by its invariant sets:

**Corollary 5.1.10** If \(C \subseteq B(X)\) is a countably generated sub-\(\sigma\)-algebra of the Borel sets of \(X\), then there exists a unique smooth equivalence relation \(\rho_C\) on \(X\) with \(C = INV (B(X), \rho_C)\).

### 5.1.2 Operations on Smooth Relations

We will study briefly operations with smooth equivalence relations. These operations will be helpful when doing constructions on stochastic relations. The interplay between smooth relations and measurable maps is further illustrated by the technique of transporting a smooth relation backwards along a measurable map.

**Lemma 5.1.11** Let \(\alpha\) be a smooth equivalence relation on the analytic space \(A\) so that \(\alpha = ker (h)\) for some measurable map \(h : A \to W\), \(W\) being an analytic space. Define for the Polish space \(X\) and the Borel map \(f : X \to A\) on \(X\) the smooth relation \(\alpha_f := ker (h \circ f)\). If \(E \subseteq X\) is an \(\alpha_f\)-invariant Borel set, then

1. \(f [E]\) is an \(\alpha\)-invariant Borel set in \(A\),

2. \(E = f^{-1} [f [E]]\).

Consequently, the invariant Borel sets of \(\alpha_f\) are just the inverse images of the invariant Borel set of \(\alpha\) under \(f\), viz.,

\[
INV (B(X), \alpha_f) = f^{-1} [INV (B(A), \alpha)].
\]

**Proof**

1. Let \(E_0 := f^{-1} [F]\) be the inverse image of an \(\alpha\)-invariant set \(F \subseteq A\), and assume that \(x \in E_0\) with \(x \alpha_f x'\). Since \(f(x) \in F\), and since \(h(f(x)) = h(f(x'))\), we have \(f(x') \in F\), thus \(x' \in E_0\). Consequently, \(E_0\) is \(\alpha_f\)-invariant, and we have shown that

\[
INV (B(X), \alpha_f) \supseteq f^{-1} [INV (B(A), \alpha)]
\]

holds.

2. Let \(E \in INV (B(X), \alpha_f)\), then we assert that \(E' := f [E] \in INV (B(A), \alpha)\). Since \(E\) is \(\alpha_f\)-invariant, \(f [E]\) is \(\alpha\)-invariant by construction. The hard part is showing that \(E'\) is a Borel set. First it is clear that \(E'\) is an analytic set, because it is the image of a Borel set under a Borel map. We claim that \(f [X \setminus E] = A \setminus f [E]\). From this we may conclude that \(E'\) is also co-analytic, thus is a Borel set by Souslin's famous theorem(Theorem A.2.5).

We first repeat the argumentation in the proof of Lemma 5.1.8 in showing that \(f [X \setminus E] \subseteq A \setminus f [E]\) holds: Suppose \(a \in f [X \setminus E]\), then we can find \(x \in X \setminus E\) with \(a = f(x)\). If \(a\) would be a member of \(f [E]\), we could find \(x' \in E\) with \(a = f(x')\). Since \(x \alpha_f x'\) and since \(E\) is \(\alpha_f\)-invariant, we would find \(x \in E\), contradicting the choice of \(x\). This establishes the desired equality and shows that \(E'\) is in fact a Borel set. But we can say more: \(E = f^{-1} [f [E]]\) will be shown to hold. Let \(x \in f^{-1} [f [E]]\), thus \(f(x) \in f [E]\), hence
Lemma 5.1.12

Assume $B$ is a Polish space, and let $\psi_1, \psi_2 : B \to Y$ be surjective Borel maps. Assume further that $\ell([\psi_1] || [\psi_2])$ is smooth. Let

$$C := \{ C \in \mathcal{B}(Y) \mid \psi_1^{-1}[C] = \psi_2^{-1}[C] \}$$

be the $\sigma$-algebra of events common to both $\psi_1$ and $\psi_2$. Then these common events are exactly the $\ell([\psi_1] || [\psi_2])$-invariant Borel sets, thus $C = INV(\mathcal{B}(Y), \ell([\psi_1] || [\psi_2]))$.

Proof 1. It is not difficult to see that if $C \in C$ is a common event then $y \in C$ and $(y, y') \in \ell([\psi_1] || [\psi_2])$ together imply $y' \in C$. This is so since $(y, y') \in \ell([\psi_1] || [\psi_2])$ implies that there exist $y_0, \ldots, y_n \in Y$ with $y_0 = y, y_n = y'$ and $(y_i, y_{i+1}) \in [\psi_1] || [\psi_2]$ for $1 \leq i \leq n - 1$. This yields

$$C \subseteq INV(\mathcal{B}(Y), \ell([\psi_1] || [\psi_2]))$$

in terms of the $\sigma$-algebras involved.

2. Now let

$$C \in INV(\mathcal{B}(Y), \ell([\psi_1] || [\psi_2])) = \eta_{\ell([\psi_1] || [\psi_2])}^{-1}[\mathcal{B}(Y)/\ell([\psi_1] || [\psi_2])],$$

the latter equality holding by Proposition 5.1.9. Thus we can find $D \in \mathcal{B}(Y)/\ell([\psi_1] || [\psi_2])$ with $C = \eta_{\ell([\psi_1] || [\psi_2])}^{-1}[D]$. Consequently,

$$\psi_1^{-1}[C] = \psi_2^{-1}[C] \iff \left( \eta_{\ell([\psi_1] || [\psi_2])} \circ \psi_1 \right)^{-1}[D] = \left( \eta_{\ell([\psi_1] || [\psi_2])} \circ \psi_2 \right)^{-1}[D].$$

Since by construction $\eta_{\ell([\psi_1] || [\psi_2])} \circ \psi_1 = \eta_{\ell([\psi_1] || [\psi_2])} \circ \psi_2$, the latter equality follows. Hence $C$ is a common event, establishing the non-trivial inclusion. \qed
The invariant sets leave a trace on the product: consider for an invariant set $P$ all pairs $(x, x') \in \alpha$ with $x \in P$. Properties on this trace and an additional property as far of measures restricted to these sets are recorded here.

**Lemma 5.1.13** Let $\alpha$ be a smooth equivalence relation on the analytic space $X$. Then

1. If $P \in INV (\mathcal{B}(X), \alpha)$, then $(P \times X) \cap \alpha = (X \times P) \cap \alpha = (P \times P) \cap \alpha$
2. $\otimes [X, \alpha] := \{(P \times X) \cap \alpha \mid P \in INV (\mathcal{B}(X), \alpha)\}$ is a $\sigma$-algebra on $\alpha$.
3. If $\mu \in \mathcal{S} (X, INV (\mathcal{B}(X), \alpha))$ is a sub-probability measure on the $\alpha$-invariant Borel sets of $X$, then $\mu^* ((P \times X) \cap \alpha) := \mu (P)$ defines a sub-probability measure on $\otimes [X, \alpha]$.

**Proof** 1. Part 1 is proved by a direct calculation, and part 2 is established by checking the defining properties of a $\sigma$-algebra.

2. For proving part 3 it is observed that $\mu^*$ is well defined, since $(P_1 \times X) \cap \alpha = (P_2 \times X) \cap \alpha$ implies $P_1 = P_2$ for the $\alpha$-invariant sets $P_1, P_2$. The properties of a finite measure are then easily established.

When dealing with bisimulations, we will have to have a look at properties of a smooth relation as a subset of the Cartesian product. Here Lemma 5.1.13 will come in handy.

Factoring a factor space through a smooth relation will not really bring new structural information: we will show that the iterated factor space is isomorphic to a factor space that can be obtained from a relation on the base space. This will be an occasion to introduce a kind of multiplicative operation on relations for later use. Then we will show that other operations such as sums, intersections and countable products of smooth relations will also lead to smooth relations.

Assume that $\rho$ is a smooth equivalence relation on the analytic space $X$, and that $\tau$ is a smooth equivalence on $X/\rho$. Define for $x, x' \in X$

$$x \cdot (\tau \bullet \rho) x' \iff [x]_\rho \tau [x']_\rho.$$  

**Proposition 5.1.14** The equivalence relation $\tau \bullet \rho$ is smooth, and the analytic spaces $X/\tau \bullet \rho$ and $(X/\rho)/\tau$ are Borel isomorphic.

**Proof** 0. Since $\tau$ is smooth, there exists a sequence $(A_n)_{n \in \mathbb{N}}$ of Borel sets $A_n \subseteq X/\rho$ which determines it. Then $(\eta_\rho^{-1} \{A_n\}_{n \in \mathbb{N}})$ determines $\tau \bullet \rho$. Its members are by construction Borel sets in $X$.

1. Define

$$g_{\rho, \tau} ([x]_{\tau \bullet \rho}) := [x]_{\rho \tau},$$

then $g_{\rho, \tau} : X/\tau \bullet \rho \to (X/\rho)/\tau$ is well-defined and turns out to be a bijection. The construction shows that $g_{\rho, \tau} \circ \eta_{\tau \bullet \rho} = \eta_{\tau} \circ \eta_{\rho}$ holds, putting $h_{\rho, \tau} := g_{\rho, \tau}^{-1}$, we see that $\eta_{\tau \bullet \rho} = h_{\rho, \tau} \circ \eta_{\tau} \circ \eta_{\rho}$. This is also noted for later use.

2. Let $E \subseteq (X/\rho)/\tau$ be a Borel set, we need to show that $g_{\rho, \tau}^{-1} [E]$ is a Borel set in $X/\tau \bullet \rho$, equivalently, that $E_0 := \eta_{\tau \bullet \rho}^{-1} [g_{\rho, \tau}^{-1} [E]]$ is an $\tau \bullet \rho$-invariant Borel set in $X$ by Proposition 5.1.9. But $E_0 = (\eta_{\rho} \circ \eta_{\tau})^{-1} [E]$, so that $E_0$ is a Borel set by the measurability of the projections, and this set is clearly $\tau \bullet \rho$-invariant. Thus we get the measurability of $E_0$ again from Lemma 5.1.9.
3. Let $F \subseteq X/\tau \mathbin{•} \rho$ be a Borel set, hence $F_0 := \eta_{\tau\rho}^{-1}[F]$ is a Borel set in $X$, thus there exists a Borel set $F_1 \subseteq (X/\rho)/\tau$, such that $F_0 = \eta_{\tau}^{-1}[\eta_{\tau\rho}^{-1}[F_1]]$ since $F_0$ is $\rho$-invariant. Hence $F_1 = h_{\rho\tau}^{-1}[F]$, so $h_{\rho\tau}$ is measurable, establishing the claim. \( \dashv \)

The definition of $\tau \mathbin{•} \rho$ translates a partition of the $\rho$-classes into a partition of the base set. The generated $\tau \mathbin{•} \rho$-partition is coarser than the $\rho$-partition, since $\rho \subseteq \tau \mathbin{•} \rho$. The converse holds as well: whenever we have two partitions coming from smooth equivalence relations, we may find a factor in terms of $\mathbin{•}$ relating them to each other. Surprisingly, this result about the containment of equivalence relations may be used in Section 5.1.3 to show that two completely unrelated equivalences may give rise to some sort of confluence.

**Corollary 5.1.15** The following conditions are equivalent for smooth equivalence relations $\rho$ and $\sigma$ on $X$:

1. $\rho \subseteq \sigma$,

2. there exists a smooth equivalence relation $\theta$ on $X/\rho$ such that $\sigma = \theta \mathbin{•} \rho$.

**Proof** 1. The direction $2 \Rightarrow 1$ is trivial, so we are left with the proof for $1 \Rightarrow 2$.

2. Define $f \left( [x]_{\rho} \right) := [x]_{\sigma}$ for $[x]_{\rho} \in X/\rho$ then $f : X/\rho \rightarrow X/\sigma$ is well defined and surjective.

We claim that $f$ is $B(X/\rho)$-$B(X/\rho)$-measurable. In fact, we need to show that $f^{-1}[D] \in \mathcal{INV}(B(X), \rho)$ whenever $D \in B(X/\rho)$. Since by Corollary 5.1.9

$$D \in B(X/\rho) \Leftrightarrow \eta_{\sigma}^{-1}[D] \in \mathcal{INV}(B(X), \sigma),$$

the claim follows from $\eta_{\rho}^{-1}[f^{-1}[D]] = \eta_{\sigma}^{-1}[D]$ and from $\mathcal{INV}(B(X), \sigma) \subseteq \mathcal{INV}(B(X), \rho)$, which is inferred from $\rho \subseteq \sigma$ by a trivial calculation. Since $\ker(f)$ is smooth by Lemma 5.1.8, we can set $\theta := \ker(f)$ and are done. \( \dashv \)

We turn now to sums and products. Knowing that the smooth equivalence relations are closed under the sum operation is helpful when discussing congruences on the sum of stochastic relations. It is clear that the sum and the product of a countable number of analytic spaces is analytic again.

**Lemma 5.1.16** Let $X$ and $Y$ be analytic spaces with smooth equivalence relations $\alpha$ resp. $\beta$. Then $\alpha + \beta := \alpha \cup \beta$ is a smooth equivalence relation on $X + Y$.

**Proof** If $(A_n)_{n \in \mathbb{N}}$ and $(B_n)_{n \in \mathbb{N}}$ determine $\alpha$ resp. $\beta$, then the countable set of Borel sets $\{A_n + B_m \mid n, m \in \mathbb{N}\}$ determines $\alpha + \beta$. \( \dashv \)

Smooth equivalence relations are closed under intersections and under (countably infinite) products.

**Lemma 5.1.17** If $\rho, \rho'$ are smooth equivalence relations on the analytic space $X$, then $\rho \cap \rho'$ is smooth, and $\mathcal{INV}(B(X), \rho \cap \rho') = \sigma(\mathcal{INV}(B(X), \rho) \cup \mathcal{INV}(B(X), \rho'))$.

**Proof** Assume that $(A_n)_{n \in \mathbb{N}}$ and $(A'_n)_{n \in \mathbb{N}}$ determine $\rho$ resp. $\rho'$. Then the sequence $(A_n \cap A'_m)_{n, m \in \mathbb{N}}$ determines $\rho \cap \rho'$. The representation for the invariant Borel sets for $\rho \cap \rho'$ follows then easily from Proposition 5.1.9. \( \dashv \)

The closure under countably infinite products will resort to the construction of the Borel sets on an infinite product through cylinder sets (cp. Section A.1).
Lemma 5.1.18 Assume that \((X_n)_{n \in \mathbb{N}}\) is a sequence of analytic spaces, and let \(\rho_n\) be a smooth equivalence relation on \(X_n\) for each \(n \in \mathbb{N}\). Define

\[(a_n)_{n \in \mathbb{N}} \sim (a'_n)_{n \in \mathbb{N}} \iff \forall n \in \mathbb{N}: a_n \rho_n a'_n.\]

Then

1. \(\times_{n \in \mathbb{N}} \rho_n\) is a smooth equivalence relation on \(\prod_{n \in \mathbb{N}} X_n\).
2. \(\mathcal{I}N(V(B(\prod_{n \in \mathbb{N}} X_n), \times_{n \in \mathbb{N}} \rho_n)) = \bigotimes_{n \in \mathbb{N}} \mathcal{I}N(V(B(X_n), \rho_n))\).

Proof 1. Abbreviate the equivalence relation \(\times_{n \in \mathbb{N}} \rho_n\) by \(\rho^\infty\). Assume that \(\rho_n\) is determined by the sequence \((Z_{n,m})_{m \in \mathbb{N}}\) of Borel sets \(Z_{n,m} \subseteq X_n\). Put

\[W_{n_1,\ldots,n_k} := Z_{1,n_1} \times \cdots \times Z_{k,n_k} \times \prod_{j > k} X_j,\]

and let \(\mathcal{E}\) be the set of all finite sequences over \(\mathbb{N}\), which is countable. Then it is easy to see that

\[(a_n)_{n \in \mathbb{N}} \rho^\infty (a'_n)_{n \in \mathbb{N}} \iff \forall \langle n_1, \ldots, n_k \rangle \in \mathcal{E}: \langle a_1, \ldots, a_k \rangle \in W_{n_1,\ldots,n_k} \iff \langle a'_1, \ldots, a'_k \rangle \in W_{n_1,\ldots,n_k} .\]

This implies that \(\rho^\infty\) is generated through a countable family of Borel sets.

2. Since for each index \(e \in \mathcal{E}\) the set \(W_e\) is a \(\rho^\infty\)-invariant Borel set which is comprised from \(\rho_n\)-invariant factors, we have

\[\mathcal{I}N(V(B(\prod_{n \in \mathbb{N}} X_n), \rho^\infty)) \subseteq \bigotimes_{n \in \mathbb{N}} \mathcal{I}N(V(B(X_n), \rho_n)),\]

on the other hand,

\[\bigotimes_{n \in \mathbb{N}} \mathcal{I}N(V(B(X_n), \rho_n))\]

is generated by cylinder sets of the form

\[B_1 \times \cdots \times B_n \times \prod_{j > n} X_j,\]

which are \(\rho^\infty\)-invariant. This implies the other inclusion. \(\dashv\)

A finite version is here available as well: the product of two smooth equivalence relations is smooth again, and the invariant Borel sets for the product are just the product of the Borel sets for the factors.

5.1.3 A Confluence Property

We will establish a confluence property that will be helpful for understanding the relationship between congruences and bisimulations, at least in a special situation. Then this confluence property will be used in a crucial way for establishing that bisimilar relations have isomorphic factors. Quite apart from this, it gives some insight into the
manipulation of relations, and it indicates a rather surprising connection to selections of set-valued maps.

An equivalence relation \( \rho \) on a set \( X \) can be viewed as a set-valued map \( x \mapsto \lfloor x \rfloor_\rho \) that assigns element \( x \in X \) its equivalence class \( \lfloor x \rfloor_\rho \), hence a particular non-empty set. The existence of a measurable selector \( f \) for this set-valued map \( \lfloor \cdot \rfloor_\rho \) implies that \( \rho \) is smooth, provided we know that \( X/\rho \) is an analytic space. This is so since

\[
\begin{align*}
x \rho x' & \iff \lfloor x \rfloor_\rho = \lfloor x' \rfloor_\rho \\
& \iff \exists f(x) = f(x'),
\end{align*}
\]

so that \( \rho = \ker(f) \), hence Lemma 5.1.6 applies.

This is essentially the outline for the proof of the confluence property that relates two smooth relations on a compact metric space.

**Proposition 5.1.19** Let \( T \) be a compact metric space, and assume that \( \rho = \ker(\phi), \sigma = \ker(\psi) \) for some continuous maps \( \phi : T \to N \) and \( \psi : T \to N' \) with metric spaces \( N, N' \). There exist smooth equivalence relations \( \theta \) on \( T/\sigma \) and \( \theta' \) on \( T/\rho \) such that

1. \( \theta' \circ \rho = \theta \circ \sigma \),
2. \( \theta \) and \( \theta' \) are minimal: if \( \theta_0' \circ \rho = \theta_0 \circ \sigma \), for smooth equivalence relations \( \theta_0 \) on \( T/\sigma \) and \( \theta_0' \) on \( T/\rho \), then \( \theta \subseteq \theta_0 \) and \( \theta' \subseteq \theta_0' \).

The diagram visualizes this claim and suggests the characterization as a confluence property.

```
\[
\begin{array}{c}
\bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\end{array}
\]
```

Note that the universal relations on the respective factor spaces would satisfy the first condition. The statement is then somewhat trivial, so minimality will make sure that we may apply it in a sensible way.

The proof will be broken into several parts. Because of Corollary 5.1.15 we will first find a smooth equivalence relation \( \theta \) on \( T/\sigma \) such that \( \rho \subseteq \theta \circ \sigma \) holds. We will assume through the end of the proof of Proposition 5.1.19 that \( T \) is a compact metric space, and that \( \rho = \ker(\phi), \sigma = \ker(\psi) \).

**Claim 1** \( T/\sigma \) is a compact metric space when endowed with the final topology for \( \eta_\sigma \).

**Proof** Let \( d' \) be the metric on \( N' \), and put for \( t, t' \in T \)

\[
D([t]_\sigma, [t']_\sigma) := d'(\psi(t), \psi(t')),
\]

then \( D \) is a metric on \( T/\sigma \) (since \( \sigma = \ker(\psi) \) by assumption). Let \( T \) be the topology on \( T/\sigma \) induced by \( \eta_\sigma \), then a set \( G \subseteq T/\sigma \) which is \( D \)-open is also \( T \)-open. This follows easily.
from the continuity of \( \psi \). Conversely, let \( F \) be \( T \)-closed, and assume that \( ([t_n]_\sigma)_{n \in \mathbb{N}} \) be a sequence in \( F \) such that \( D([t_n]_\sigma, [t]_\sigma) \to 0 \), as \( n \to \infty \). Select an arbitrary \( x_n \in [t_n]_\sigma \), thus \( x_n \in \eta_\sigma^{-1} [F] \), the latter is a closed, hence compact set. Thus we can find a convergent subsequence (which we take w.l.g. the sequence itself), so that there exists \( x^* \in \eta_\sigma^{-1} [F] \) with \( x_n \to x^* \). By the continuity of \( \psi \) we may conclude \( \psi(x_n) \to \psi(x^*) \). This implies \( x^* \in [t]_\sigma \), thus \( [t]_\sigma \subseteq F \). Hence \( F \) is also metrically closed, and the topologies coincide. \( \dagger \)

Claim 1 shows among others that the Borel sets in \( T/\sigma \) come from a compact metric space. This observation will make some arguments easier. When talking about the topology on \( T/\sigma \), we refer interchangeably to the metric topology and the topology induced by the canonical projection.

**Claim 2** Put

\[
\zeta := \{ \langle s, s' \rangle \mid s, s' \in T/\sigma, s \times s' \cap \rho \neq \emptyset \}.
\]

Then \( \zeta \subseteq (T/\sigma)^2 \) is reflexive, symmetric, and a closed subset of \((T/\sigma)^2\).

**Proof** Since \( \rho \) is reflexive and symmetric, \( \zeta \) is. Now let \( \langle s_n, s'_n \rangle \in \zeta \) be a convergent sequence, say \( s_n \to s, s'_n \to s' \). For \( s_n \) there exists by the construction of \( \zeta \) a pair \( \langle t_n, t'_n \rangle \in \rho \) with \( t_n \in s_n, t'_n \in s'_n \). In particular, \( \phi(t_n) = \phi(t'_n) \). Compactness implies the existence of a subsequence \( (q(n))_{n \in \mathbb{N}} \) and of elements \( t, t' \) such that \( t_{q(n)} \to t, t'_{q(n)} \to t' \), as \( n \to \infty \).

Continuity implies \( \langle t, t' \rangle \in \ker(\phi) = \rho \), and \( s = [t]_\sigma, s' = [t']_\sigma \). Thus \( \zeta \) is closed. \( \dagger \)

Now define inductively the \( n \)-fold composition of \( \zeta \):

\[
\zeta^{(1)} := \zeta,
\]

\[
\zeta^{(n+1)} := \zeta^{(n)} \circ \zeta,
\]

where \( \circ \) denotes the usual relational composition.

The following properties are easily established through a compactness argument using induction on \( n \):

**Claim 3** For each \( n \in \mathbb{N} \)

1. \( \zeta^{(n)} \subseteq T/\sigma \) is closed,

2. if \( C \subseteq T/\sigma \) is compact, then the set

\[
\exists \zeta^{(n)}(C) = \{ s \in T/\sigma \mid \exists s' \in C : \langle s, s' \rangle \in \zeta^{(n)} \}
\]

is closed.

**Claim 4** The transitive closure \( \theta \) of \( \zeta \) is a smooth equivalence relation on \( T/\sigma \).

**Proof** It is clear from the properties of \( \zeta \) that \( \theta \) is an equivalence relation. Smoothness needs to be shown, and we will exhibit a Borel measurable map into a Polish space with \( \theta \) as its kernel. Write \( \theta \) as

\[
\theta = \bigcup_{n \in \mathbb{N}} \zeta^{(n)},
\]

and let \( C \subseteq T/\sigma \) be a compact set, then

\[
\exists \theta(C) = \{ s \in T/\sigma \mid \exists s' \in C : \langle s, s' \rangle \in \theta \}
\]
is a measurable subset of $T/\sigma$ by Claim 3. By Proposition A.2.7 we can find a measurable selector $w$ for the set valued map $s \mapsto \eta_\rho(s)$, hence a Borel measurable map $w : T/\sigma \to T/\sigma$ such that $w(s) \in [s]_\rho$ for each $s \in T/\sigma$. Now we know that $T/\sigma$ is a Polish space, and from $\theta = \ker(w)$ we infer that $\theta$ is smooth. \qed

We are in a position now to establish Proposition 5.1.19.

**Proof** (of Proposition 5.1.19). 1. It is sufficient for the first part to establish $\rho \subseteq \theta \cdot \sigma$. In fact, let $t \rho t'$, then $[t]_\sigma \times [t']_\sigma \cap \rho \neq \emptyset$. This implies $\langle [t]_\sigma, [t']_\sigma \rangle \in \zeta \subseteq \theta$, which in turn establishes the inclusion and hence the Proposition.

2. Now assume that $\theta_0' \cdot \rho = \theta_0 \cdot \sigma$, for smooth equivalence relations $\theta_0$ on $T/\sigma$ and $\theta_0'$ on $T/\rho$ hold, thus from Corollary 5.1.15 we infer $\sigma \subseteq \theta_0' \cdot \rho$ and $\rho \subseteq \theta_0 \cdot \sigma$. In order to establish $\theta \subseteq \theta_0$ it is enough to show that $\zeta \subseteq \theta_0$. But if $\langle s, s' \rangle \in \zeta$, we know that $s \times s' \cap \theta_0 \cdot \sigma \neq \emptyset$, thus $\langle t, t' \rangle \in \theta_0 \cdot \sigma$ for some $t \in s = [t]_\sigma, t' \in s' = [t']_\sigma$. Consequently, $s \theta_0 s'$ holds. Interchanging the roles of $\rho$ and $\sigma$ establishes that $\theta' \subseteq \theta_0' \cdot \rho$ also holds. \qed

For later use we record a property of $\theta$-invariant Borel sets that characterizes these sets in terms of the equivalence relations from which $\theta$ is constructed. It gives an easy criterion on invariance and indicates that the relation $\theta$ will have some use in the discussions to follow.

**Lemma 5.1.20** Under the assumptions of Proposition 5.1.19, let $D \subseteq T/\sigma$ be a Borel set, where $\theta$ is defined as in Claim 4 as the equivalence relation generated through $\{\langle s, s' \rangle \mid s, s' \in T/\sigma, s \times s' \cap \rho \neq \emptyset\}$. Then these conditions are equivalent:

1. $D$ is $\theta$-invariant,

2. $\eta^{-1}_\sigma[D] \in INV(B(T), \sigma) \cap INV(B(T), \rho)$.

**Proof** 1. $\Rightarrow$ 2: We know from Lemma 5.1.8 that $\eta^{-1}_\sigma[D] \in INV(B(T), \sigma)$, because $D \subseteq T/\sigma$ is a Borel set. Now let $t \in \eta^{-1}_\sigma[D]$ with $t \rho t'$. Then $[t]_\sigma \in D$ and $[t]_\sigma \times [t']_\sigma \cap \rho \neq \emptyset$, thus $\langle [t]_\sigma, [t']_\sigma \rangle \in \theta$, and, since $D$ is $\theta$-invariant, $t' \in \eta^{-1}_\sigma[D]$. Hence $\eta^{-1}_\sigma[D]$ is also $\rho$-invariant.

2. The implication $2 \Rightarrow 1$ is established through a routine argument using induction accounting for the construction of $\theta$. \qed

### 5.1.4 Spawning

We will construct a way of transporting a smooth equivalence relation along a map between equivalence classes. Such a map will be employed as a bridge between the relations, in particular we will transport vital properties along it. These properties will become apparent later, when we develop criteria for the bisimilarity of stochastic relations. Quite independent of this particular application the concept of spawning yields some insight into the nature of smooth equivalence relation in terms of the Borel structure on their factor spaces.

As a preparation for the definition of how two smooth relations relate to each other we will have a quick look at how the atoms of a countably generated $\sigma$-algebra are characterized through the generators.
\section{5.1 Smooth Equivalence Relations}

**Lemma 5.1.21** Let \( \mathcal{E} = \sigma(\{E_n | n \in \mathbb{N}\}) \) be a countably generated \( \sigma \)-algebra over a set \( E \). Define \( A^1 := A, A^0 := E \setminus A \) for \( A \subseteq E \), and put

\[
E(\alpha) := \prod_{n \in \mathbb{N}} E_\alpha(n)
\]

for \( \alpha \in \{0,1\}^\mathbb{N} \). Then there exists \( F \subseteq \{0,1\}^\mathbb{N} \) such that \( \{E(\alpha) | \alpha \in F\} \) are exactly the atoms of \( \mathcal{E} \).

**Proof** 0. The proof follows essentially \cite[Proposition 3.1.15]{88}.

We first note that as in the proof of Lemma 5.1.3,

\[
\mathcal{G}_{a,b} := \{A \subseteq E | a \in A \Leftrightarrow b \in A\}
\]

is a \( \sigma \)-algebra. It is clear that the sets \( E(\alpha) \) are elements of \( \mathcal{E} \). Define \( F \) as the set of all indices \( \alpha \) for which \( E(\alpha) \neq \emptyset \).

1. Now let \( \alpha \in F \), then \( E(\alpha) \) is an atom of \( \mathcal{E} \). Suppose it is not, then there exists \( B \in \mathcal{E} \) with \( \emptyset \neq B \subseteq E(\alpha) \), so we can pick \( a,b \in E(\alpha) \) with \( a \in B \) and \( b \notin B \). From part 0. we infer that there exists an index \( m \in \mathbb{N} \) such that \( E_m \) contains exactly one of \( a,b \). On the other hand, we see that the construction implies \( [a \in E_n \Leftrightarrow b \in E_n] \) for each \( n \in \mathbb{N} \). This is a contradiction, so \( E(\alpha) \) is an atom of \( \mathcal{E} \).

2. Let \( A \) be an atom of \( \mathcal{E} \), and put

\[
h(A) := \bigcap\{E_n | A \subseteq E_n\} \cap \bigcap\{E \setminus E_n | A \cap E_n = \emptyset\},
\]

then there exists \( \alpha \in \{0,1\}^\mathbb{N} \) such that \( h(A) = E(\alpha) \). Now \( A \subseteq h(A) \) is immediate. Because \( h(A) \) is an atom as well, we see \( A = h(A) = E(\alpha) \).

We are ready for a technical definition that permits stating how a smooth equivalence relation is transported through a map between classes in such a way that important properties are maintained. We call this spawning. This definition of spawning is at present on the level of equivalence relations, it will be later extended to incorporate congruences.

**Definition 5.1.22** Let \( \alpha \) and \( \beta \) be smooth equivalence relations on the analytic spaces \( X \) resp. \( Y \), and assume that \( \Upsilon : X/\alpha \rightarrow Y/\beta \) is a map between the equivalence classes. We say that \( \alpha \) spawns \( \beta \) via \( (\Upsilon, A_0) \) iff \( A_0 \) is a countable generator of \( \text{INV}(B(X), \alpha) \) such that

1. \( A_0 \) is closed under finite intersections,
2. \( \{\Upsilon_A | A \in A_0\} \) is a generator of \( \text{INV}(B(Y), \beta) \), where \( \Upsilon_A := \bigcup\{\Upsilon([x]_\alpha) | x \in A\} \),
3. \([x_1]_\alpha = [x_2]_\alpha\) implies the equality of

\[
\bigcap\{\Upsilon_A | A \in A_0, x_1 \in A\} \cap \bigcap\{X' \setminus \Upsilon_A | A \in A_0, x_1 \notin A\}
\]

and

\[
\bigcap\{\Upsilon_A | A \in A_0, x_2 \in A\} \cap \bigcap\{X' \setminus \Upsilon_A | A \in A_0, x_2 \notin A\}.
\]
Thus if \( \alpha \) spawns \( \beta \), then the measurable structure induced by \( \alpha \) on \( X \) is all we need for constructing the measurable structure induced by \( \beta \) on \( Y \): the map \( \Upsilon \) can be made to carry over the generator \( \mathcal{A}_0 \) from \( \mathcal{I} \mathcal{N} \mathcal{V} (\mathcal{B}(X), \alpha) \) to \( \mathcal{I} \mathcal{N} \mathcal{V} (\mathcal{B}(Y), \beta) \) and — in the light of Lemma 5.1.21 — to transport the atoms from one \( \sigma \)-algebra to the other. This is of particular interest since the atoms are just the equivalence classes. Hence \( \alpha \) together with \( \Upsilon \) and the generator \( \mathcal{A}_0 \) is all we may care to know or to learn about \( \beta \). The first condition reflects a measure-theoretic precaution: we will need to make sure e.g. in the construction of the direct sum of stochastic relations that measures are uniquely determined by their values on a generators. This, however, can best be guaranteed if the generator is stable against taking finite intersections, as witnessed by the \( \pi \)-\( \lambda \)-Theorem A.1.1. Note that \( \Upsilon \mathcal{A}_1 \cap \mathcal{A}_2 = \Upsilon \mathcal{A}_1 \cap \Upsilon \mathcal{A}_2 \) also holds, so that closedness under intersections is inherited through \( \Upsilon \).

Whenever we have two smooth equivalence relations such that one spawns the other we obtain on the sum of the underlying spaces a unique smooth relation the traces of which on the summands are just the given relations. Since we will introduce later on the sum of two relations, this effect will be studied now carefully.

**Lemma 5.1.23** Let \( \alpha \) and \( \beta \) be smooth equivalence relations on the analytic spaces \( X \) resp. \( Y \).

1. If \( \alpha \) spawns \( \beta \) via \( (\Upsilon \mathcal{A}_0) \), then there exists exactly one smooth equivalence relation \( \gamma \) on \( X + Y \) such that \([x]_\gamma \cap X = [x]_\alpha \) and \([x]_\gamma \cap Y = \Upsilon([x]_\alpha) \) holds for all \( x \in X \).

2. If in addition \( \beta \) spawns \( \alpha \) via \( (\Theta, \mathcal{B}_0) \), then we have for all \( x \in X, y \in Y \) the equivalence

\[
[y]_\beta = \Upsilon([x]_\alpha) \iff [x]_\alpha = \Theta([y]_\beta).
\]

**Proof 1.** In order to establish property 1, we consider the equivalence relation \( \gamma \) generated from \( \{ \mathcal{A}_n + \Upsilon \mathcal{A}_n \mid n \in \mathbb{N} \} \) with \( \mathcal{A}_0 = \{ \mathcal{A}_n \mid n \in \mathbb{N} \} \). Relation \( \gamma \) is evidently smooth, and it is uniquely determined through \( \alpha \) and \( \beta \). Let \( x \in X \), then we can find \( y \in Y \) with \( \Upsilon([x]_\alpha) = [y]_\beta \), since \( \Upsilon \) maps \( X/\alpha \) to \( Y/\beta \). It is easy to see that

\[
\Upsilon([x]_\alpha) = \bigcap \{ \Upsilon \mathcal{A}_n \mid \mathcal{A}_n \in \mathcal{A}_0, x \in \mathcal{A}_n \} \cap \bigcap \{ Y \setminus \Upsilon \mathcal{A}_n, \mathcal{A}_n \in \mathcal{A}_0, x \notin \mathcal{A}_n \}.
\]

This establishes part 1.

2. The claim in part 2 follows from the observation that \( \beta \) is the relation on \( Y \) which is generated from \( \{ \Theta B_n \mid n \in \mathbb{N} \} \), and that \( \alpha \) is the relation on \( X \) which is generated from \( \{ \Theta B_n \mid n \in \mathbb{N} \} \) where \( \mathcal{B}_0 = \{ B_n \mid n \in \mathbb{N} \} \). \( \Box \)

The isomorphism of two factor systems is an illustration of the concept of spawning.

**Proposition 5.1.24** Let \( T, T' \) be analytic spaces with smooth equivalence relations \( \rho \) resp. \( \rho' \). Assume that \( \Upsilon : T/\rho \rightarrow T'/\rho' \) is a Borel isomorphism, and let \( \mathcal{A} \) be a countable generator of \( \mathcal{I} \mathcal{N} \mathcal{V} (\mathcal{B}(X), \rho) \) which is closed under finite intersections. Then \( \rho \) spawns \( \rho' \) via \( (\Upsilon, \mathcal{A}) \).

**Proof 0.** The assumption that the generator \( \mathcal{A} \) is closed under finite intersections is easily met: take an arbitrary countable generator \( \mathcal{A}_0 \), then

\[
\{ \bigcap \mathcal{F} \mid \mathcal{F} \subseteq \mathcal{A}_0 \text{ is finite} \}
\]

130
is a countable generator which is closed under finite intersections.

1. If \( A \in \mathcal{I}NV (\mathcal{B}(T), \rho) \) is a \( \rho \)-invariant Borel set, then \( \Upsilon_A \) is \( \rho' \)-invariant, and it is easily established that

\[
\Upsilon_A = \eta_{\rho}^{-1} [\Upsilon [\eta_{\rho} [A]]]
\]

holds. From Proposition 5.1.9 we infer that \( \mathcal{C}_1 \) is a generator for \( \mathcal{B}(T/\rho) \), where

\[
\mathcal{C}_1 := \eta_{\rho} \left[ \mathcal{I}NV (\mathcal{B}(T), \rho) \right].
\]

Consequently, \( \{ \Upsilon_A \mid A \in A \} \) is a generator for the \( \rho' \)-invariant Borel sets \( \mathcal{I}NV (\mathcal{B}(T'), \rho') \), because we may conclude

\[
\mathcal{I}NV (\mathcal{B}(T'), \rho') = \eta_{\rho'}^{-1} \left[ \mathcal{B}(T'/\rho') \right] \quad \text{(by Lemma 5.1.8)}
\]

\[
= \eta_{\rho'}^{-1} \left[ \Upsilon [\mathcal{B}(T'/\rho')] \right] \quad \text{(since \( \Upsilon \) is a Borel isomorphism)}
\]

\[
= \sigma \left( \{ \eta_{\rho'}^{-1} [\Upsilon [C]] \mid C \in \mathcal{C}_1 \} \right) \quad \text{(by construction of \( \mathcal{C}_1 \))}
\]

\[
= \sigma \left( \{ \Upsilon_A \mid A \in \mathcal{I}NV (\mathcal{B}(T), \rho) \} \right)
\]

\[
= \sigma \left( \{ \Upsilon_A \mid A \in \mathcal{A} \} \right) \quad \text{(since \( \mathcal{A} \) generates \( \mathcal{I}NV (\mathcal{B}(T), \rho) \)).}
\]

2. As in the proof of Proposition 5.1.9 one shows that

\[
\Upsilon \left( \left[t\right]_\rho \right) = \bigcap \{ \eta_{\rho'} [\Upsilon_A] \mid A \in \mathcal{A}, t \in A \} \cap \bigcap \{ \eta_{\rho'} [T' \setminus \Upsilon_A] \mid A \in \mathcal{A}, t \notin A \}.
\]

This settles the condition of well-defined atoms and concludes the proof. \( \dashv \)

The proof is technically a bit laborious. The statement, however, will be most useful in permitting us to show that stochastic relations are bisimilar, provided they have isomorphic factors. Working with isomorphisms alone for characterizing bisimilarity may be too strong a condition. In the application to modal logic in section 6.1 we will see that the equivalence relation which is induced on states through having the same logic satisfies the condition on spawning, but it is far from clear in this case whether or not the corresponding factor spaces are Borel isomorphic. Consequently, it seems to be worthwhile to work with the weaker condition.

### 5.2 Factoring

Observing a stochastic relation \( \mathcal{K} = (X, Y, K) \), elements with equivalent behavior are identified. This leads to a pair \((\alpha, \beta)\) of equivalence relations on the inputs \( X \) resp. the outputs \( Y \) with the idea that equivalent inputs lead to equivalent outputs. While equivalent inputs can be described directly through \( \alpha \), the equivalence of outputs requires a description on the level of measurable sets. This leads then naturally to the notion of a congruence, which will be defined in this section. We will investigate congruent systems and present some technical properties that shed some light on the underlying invariant sets. This in turn will help us to investigate properties of specific congruences.

#### 5.2.1 Congruences

Think of a stochastic relation \( \mathcal{K} = (X, Y, K) \) as a model that relates inputs and outputs, and assume that there are equivalences \( \alpha \) and \( \beta \) on inputs resp. outputs. Two inputs
\(x, x' \in X\) cannot be distinguished through \(\alpha\) iff they are \(\alpha\)-equivalent. It is less intuitive to describe distinguishing outputs through \(\beta\), in particular when we do not have a handle on specific outputs but rather on sets of them. We argue that a set \(B \subseteq Y\) cannot be distinguished through \(\beta\) iff whenever \(y \in B\) and \(y \beta y'\) we have \(y' \in B\) as well (or: it must not happen that \(y \in B\), but some \(y'\) with \(y' \beta y\) fails to be a member of \(B\)). This entails that \(B\) is invariant with respect to \(\beta\).

This consideration leads to the definition of a congruence for \(K\).

**Definition 5.2.1** A congruence \(c = (\alpha, \beta)\) for the stochastic relation \(K = (X, Y, K)\) over the analytic spaces \(X\) and \(Y\) is a pair of smooth equivalence relations \(\alpha\) on \(X\) and \(\beta\) on \(Y\) such that \(K(x)(D) = K(x')(D)\) holds whenever \(x \alpha x'\) and \(D\) is a \(\beta\)-invariant measurable subset of \(Y\).

In algebraic theories, kernels of morphisms and congruences are basically the same thing. This is also true in the present case. Denote for the morphism \(f : K_1 \to K_2\) with \(f = (\phi, \psi)\) its kernel \(\ker(f)\) by the pair \((\ker(\phi), \ker(\psi))\).

**Proposition 5.2.2** If \(f : K \to K'\) is a morphism for the stochastic relations \(K\) and \(K'\), then \(\ker(f)\) is a congruence for \(K\).

**Proof** Let \(K = (X, Y, K)\) and \(K' = (X', Y', K')\) with \(f = (\phi, \psi)\). Let \(x_1 \ker(\phi) x_2\) and \(D \subseteq Y\) be a \(\ker(\psi)\)-invariant Borel subset of \(Y\). Lemma 5.1.8 shows that \(D = \psi^{-1}[D']\) for some Borel set \(D' \subseteq Y'\). Thus

\[
K(x_1)(D) = K(x_1)(\psi^{-1}[D'])
\]

\[
= (\mathcal{S}(\psi) \circ K)(x_1)(D')
\]

\[
= (K' \circ \phi)(x_1)(D')
\]

\[
= K(\phi(x_1))(D')
\]

\[
= K(\phi(x_2))(D')
\]

\[
= K(x_2)(D),
\]

since \(f = (\phi, \psi)\) is a morphism. \(\square\)

This construction permits introducing factor objects. They will be heavily used throughout.

**Proposition 5.2.3** Let \(c = (\alpha, \beta)\) be a congruence on the stochastic relation \(K = (X, Y, K)\) with analytic spaces \(X\) and \(Y\), and define

\[
K_{\alpha, \beta}([x]_\alpha)(D) := K(x)(\eta_\beta^{-1}[D])
\]

for \(x \in X, D \in \mathcal{B}(Y/\beta),\) then

1. \(K_{\alpha, \beta} : X/\alpha \rightsquigarrow Y/\beta\) defines a stochastic relation \(K/c\) over the analytic spaces \(X/\alpha\) and \(Y/\beta\),

2. \(\eta_c := (\eta_\alpha, \eta_\beta) : K \to K/c\) is a morphism.

We call

\[
K/c := (X/\alpha, Y/\beta, K_{\alpha, \beta})
\]

the factor object (of \(K\) with respect to \(c\)).
Proof 0. We know from Proposition 5.1.7 that the factor spaces \( X/\alpha \) and \( Y/\beta \) are analytic spaces.
1. Given \( D \in \mathcal{B}(Y/\beta) \), \( \eta_\beta^{-1}[D] \) is an invariant Borel set, thus \( x \mapsto K(x)(\eta_\beta^{-1}[D]) \) does depend only on the \( \alpha \)-class of \( x \in X \). Consequently, \( K_{\alpha,\beta} \) is well-defined.
2. \( K_{\alpha,\beta} : X/\alpha \rightsquigarrow Y/\beta \) is a stochastic relation. In fact, it is plain that \( K_{\alpha,\beta}([x]_\alpha) \) is a sub-probability measure on \( \mathcal{B}(Y/\beta) \), so it remains to show that \( t \mapsto K_{\alpha,\beta}(t)(D) \) is a \( \mathcal{B}(X/\alpha) \)-measurable map for each \( D \in \mathcal{B}(Y/\beta) \). Fix such a \( D \) and a Borel set \( F \subseteq \mathbb{R} \), then
\[
F_D := \{ x \in X \mid K(x)(\eta_\beta^{-1}[D]) \in F \}
\]
is a Borel set in \( X \), and since \( \eta_\beta^{-1}[D] \) is \( \beta \)-invariant, \( F_D \) is \( \alpha \)-invariant with
\[
\{ t \in X/\alpha \mid K_{\alpha,\beta}(t)(D) \in F \} = \eta_\alpha[F_D] \in \mathcal{B}(X/\alpha)
\]
by Corollary 5.1.9. This establishes measurability.
3. The construction of \( K_{\alpha,\beta} \) yields \( K_{\alpha,\beta} \circ \eta_\alpha = \mathcal{S}(\eta_\beta) \circ K \), hence \( \eta_c \) is a morphism. \( \dashv \)
Let us see what happens if the second component \( \beta \) is the universal relation. If fact, let \( K = (X,Y,K) \) be a Polish object such that for simplicity \( K(x)(Y) = 1 \) holds for each \( x \in X \). Since we know that for the universal relation \( U_Y \) on \( Y \) the invariant Borel sets are just \( \{ \emptyset, Y \} \), (see page 118), it is clear that \( (\alpha, U_Y) \) is a congruence for an arbitrary smooth relation \( \alpha \) on \( X \). But it says only that \( K(x)(Y) = K(x')(Y) \) and \( K(x)(\emptyset) = K(x')(\emptyset) \) hold, whenever \( x \alpha x' \), so it is quite trivial.

Definition 5.2.4 Call a congruence \( c = (\alpha, \beta) \) for a stochastic relation \( K = (X,Y,K) \) non-trivial iff \( \beta \neq U_Y \).

We will have to take care of the non-triviality of congruences when investigating the problem of the bisimilarity of relations.
Restricting a stochastic relation to the invariant sets of a congruence yields a stochastic relation again; this will be of use when discussing sets of states that accept the same formula of a logic (cp. section 6.3).

Lemma 5.2.5 Let \( (\alpha, \beta) \) be a congruence for the stochastic relation \( K : X \rightsquigarrow Y \). Then
\[
K : (X, INV(\mathcal{B}(X), \alpha)) \rightsquigarrow (Y, INV(\mathcal{B}(Y), \beta))
\]
is a stochastic relation.

Proof 0. We need to establish the following: If \( B \in INV(\mathcal{B}(Y), \beta) \) is a \( \beta \)-invariant Borel set in \( Y \), and \( E \subseteq \mathbb{R}_+ \) is a Borel set in the real line, then
\[
(K(\cdot)(B))^{-1}[E] = \{ x \in X \mid K(x)(B) \in E \}
\]
is an \( \alpha \)-invariant Borel set in \( X \).
1. In fact, let \( K(x)(B) \in E \), and assume that \( x \alpha x' \). Since \( B \in INV(\mathcal{B}(Y), \beta) \), we know that \( K(x)(B) = K(x')(B) \). This establishes the assertion, since \( \{ x \in X \mid K(x)(B) \in E \} \) is a Borel set on account of \( K : (X, \mathcal{B}(X)) \rightsquigarrow (Y, \mathcal{B}(Y)) \) being a stochastic relation. \( \dashv \)
We obtain as a consequence that the integral for functions that are measurable with the respect to the invariant sets have some invariance properties.
Corollary 5.2.6 Let \((\alpha, \beta)\) be a congruence for the stochastic relation \(K : X \sim Y\). Assume furthermore that \(f : Y \to \mathbb{R}\) is bounded real-valued function which is \(\mathcal{I}NV(\mathcal{B}(Y), \beta)\)-\(\mathcal{B}(\mathbb{R})\)-measurable. Then
\[
\int_W f \, dK(x) = \int_W f \, dK(x'),
\]
whenever \(x \alpha x'\).

Proof Because \(f\) can be decomposed into a positive and a negative part, we may and do assume that \(f : Y \to \mathbb{R}_+\) holds. We know from Proposition A.1.2 that \(f\) can be approximated from below by step functions, i.e. by functions of the form 
\[
f_n := \sum_{i=0}^{k_n} q_{n,i} \cdot \chi_{A_{n,i}}
\]
with coefficients \(q_{n,i} \geq 0\) and \(A_{n,i} \in \mathcal{I}NV(\mathcal{B}(Y), \beta)\). Thus \(f_1(y) \leq f_2(y) \leq f_3(y) \ldots\), and \(f(y) = \sup_{n \in \mathbb{N}} f_n(y)\) holds for all \(y \in Y\). But then we obtain from the Bounded Convergence Theorem for \(x \alpha x'\):
\[
\int_Y f \, dK(x) = \lim_{n \to \infty} \int_Y f_n \, dK(x) = \lim_{n \to \infty} \sum_{i=0}^{k_n} q_{n,i} \cdot K(x)(A_{n,i}) = \lim_{n \to \infty} \int_Y f_n \, dK(x') = \int_Y f \, dK(x').
\]
This settles the assertion. \(\dashv\)

Remark: An alternative to the proof above observes the folklore equation
\[
\int_Y f \, dK(x) = \int_0^\infty K(x)(\{f > t\}) \, dt,
\]
where \(\{f > t\} := \{y \in Y \mid f(y) > t\}\) and \(f \geq 0\). This is a \(\beta\)-invariant Borel subset of \(Y\). Because of this set’s invariance, we see that
\[
K(x)(\{f > t\}) = K(x')(\{f > t\})
\]
holds for each \(t\), establishing the claim. —

This section has provided us with a small set of quite effective tools. Questions pertaining smooth equivalence relations will occur over and over again, so that we provide here a concise, central locus of information.

5.2.2 Isomorphism Theorems

Now fix an analytic object \(K = (X, Y, K)\), and let \(c = (\rho, \tau)\) be a congruence on \(K\). Assume that \(d = (\kappa, \lambda)\) is a congruence of \(K/c\). Define \(d \cdot c := (\kappa \bullet \rho, \lambda \bullet \tau)\).

Proposition 5.2.7 \(d \cdot c\) is a congruence on \(K\), and \(K/d \cdot c\) is isomorphic to \((K/c)/d\).
**Proof** 1. The first assertion follows from Corollary 5.2.2 together with the observation that $(\kappa \bullet \rho, \lambda \bullet \tau) = (\ker (\eta_\kappa \circ \eta_\rho), \ker (\eta_\lambda \circ \eta_\tau))$ holds.

2. Construct the Borel isomorphisms $g_{\rho, \kappa} : X/\kappa \bullet \rho \to (X/\rho)/\kappa$ and $g_{\tau, \lambda} : Y/\lambda \bullet \tau \to (Y/\tau)/\lambda$ with their respective inverses $h_{\rho, \kappa}$ and $h_{\tau, \lambda}$ as in the proof of Proposition 5.1.14. We show that the inner and the outer diagram

\[ \begin{array}{ccc}
X/\kappa \bullet \rho & \xrightarrow{g_{\rho, \kappa}} & (X/\rho)/\kappa \\
\downarrow & & \downarrow \text{\textcircled{K}}_{\rho, \tau, \lambda} \\
\mathcal{S} (Y/\lambda \bullet \tau) & \xrightarrow{\mathcal{S} (h_{\tau, \lambda})} & \mathcal{S} ((Y/\tau)/\lambda)
\end{array} \]

both commute.

3. Let $B \in \mathcal{B}((Y/\tau)/\lambda)$, a Borel set in $(Y/\tau)/\lambda$, then

\[ K_{\kappa \bullet \rho, \lambda \bullet \tau} ([x]_{\kappa \bullet \rho}) (g_{\tau, \lambda}^{-1} [B]) = K(x) (\eta_\lambda^{-1} [\eta_\kappa^{-1} [B]]) = K(x) (\eta_\lambda^{-1} [\eta_\tau^{-1} [B]]) = K_{\rho, \tau} ([x]_\rho) (\eta_\lambda^{-1} [B]) = (K_{\rho, \tau})_{\kappa, \lambda} (g_{\rho, \kappa}([x]_\rho))(B), \]

because $g_{\tau, \lambda} \circ \eta_{\lambda \bullet \tau} = \eta_\lambda \circ \eta_\tau$. Thus the outer diagram commutes. This implies that

\[ g := (g_{\rho, \kappa} : g_{\beta, \tau}) : K/d \bullet c \to (K/c)/d \]

is a morphism.

4. Suppose that $G \in \mathcal{B}(Y/\lambda \bullet \tau)$ is a Borel set, then

\[ K_{\kappa \bullet \rho, \lambda \bullet \tau} (h_{\rho, \kappa} ([x]_\rho))(G) = K_{\kappa \bullet \rho, \lambda \bullet \tau} ([x]_{\kappa \bullet \rho})(G) = K(x) (\eta_\lambda^{-1} [G]) = K_{\rho, \tau} ([x]_\rho) (\eta_\lambda^{-1} [h_{\tau, \lambda}^{-1} [G]]) = (K_{\rho, \tau})_{\kappa, \lambda} (h_{\beta, \tau}^{-1} [G]). \]

This is so since $\eta_{\lambda \bullet \tau} = h_{\tau, \lambda} \circ \eta_\lambda \circ \eta_\tau$ holds (see the proof of Proposition 5.1.14). Thus the inner diagram commutes. This implies that

\[ h := (h_{\rho, \kappa}, h_{\beta, \tau}) : (K/c)/d \to K/d \bullet c \]

is a morphism. It is plain that $h$ is left- and right inverse to $g$. ⊣

Factoring a stochastic relation with a congruence entails identifying inputs resp. outputs that have been observed as representing identical behavior. Proposition 5.2.7 says then that identifying identical behavior in observing the factor system amounts to a system that can also be obtained through a single observational step from the original system. This means that there are no arbitrary long chains of factor systems which could not have been obtained directly from the original system, or, that factoring does...
Corollary 5.2.9
Assume that \( \psi \) because \( \phi \) are well-defined. Since \( \phi \) isomorphic to a factor. A similar but slightly stronger construction for coalgebras is carried out by Rutten [77, Theorem 7.4] in the context of bisimulation relations for coalgebras. Proposition 5.2.7 and Rutten’s Theorem are not directly comparable, however, since the functor underlying the coalgebra is assumed to preserve weak pullbacks (which is no realistic assumption for stochastic relations by Corollary 4.3.7), and since the relationship between bisimulations and congruences is slightly less involved in the coalgebraic case.

Let \( (\alpha, \beta) \) and \( (\alpha', \beta') \) be pairs of equivalence relations, and define \( (\alpha, \beta) \preceq (\alpha', \beta') \) iff \( \alpha \) refines \( \alpha' \) and \( \beta \) refines \( \beta' \) simultaneously, formally:

\[
(\alpha, \beta) \preceq (\alpha', \beta') \iff \alpha \subseteq \alpha' \text{ and } \beta \subseteq \beta'
\]

It is clear that \( c \preceq d \circ c \) for each congruence \( d \).

**Proposition 5.2.8** Assume that \( f : K \to K' \) is a morphism, and let \( c \) be a congruence on \( K \) such that \( c \preceq \ker(f) \). Then there exists a unique morphism \( f_c : K/c \to K' \) with \( f = f_c \circ \eta_c \).

**Proof** 1. Let \( K = (X,Y,K), K' = (X',Y',K') \) with \( \phi : X \to X', \psi : Y \to Y' \) constituting morphism \( f \), and \( c = (\alpha, \beta) \). Because \( \alpha \subseteq \ker(\phi), \beta \subseteq \ker(\psi) \), the maps

\[
\phi_\alpha([x]_\alpha) := \phi(x), \\
\psi_\beta([y]_\beta) := \psi(y)
\]

are well-defined. Since \( \phi \) is \( B(X)-B(X') \)-measurable, and since \( B(X)/\alpha \) is the final \( \sigma \)-algebra on \( X/\alpha \) with respect to \( \eta_\alpha \), \( B(X)/\alpha-B(X') \)-measurability of \( \phi_\alpha \) is inferred. A similar argument is used for \( \psi_\beta \). Clearly, these maps are onto.

2. It remains to show that \( f_c := (\phi_\alpha, \psi_\beta) \) is a morphism. In fact, let \( D' \subseteq Y' \) be a Borel set, then

\[
K'(\phi_\alpha([x]_\alpha))(D') = K'(\phi(x))(D') = K(\phi^{-1}(D')) = K_{\alpha,\beta}([x]_\alpha)(\psi_\beta^{-1}(D')) = (\mathcal{S}(\psi_\beta) \circ K_{\alpha,\beta})([x]_\alpha)(D'),
\]

because \( \psi^{-1}(D') = \eta_\beta^{-1} \left[ \psi^{-1}(D') \right] \), and because \( (\eta_\alpha, \eta_\beta) \) is a morphism. Consequently, the equality

\[
K' \circ \phi_\alpha = \mathcal{S}(\psi_\beta) \circ K_{\alpha,\beta}
\]

has been established. Uniqueness follows, since \( \eta_c \) is an epi. \( \dashv \)

**Corollary 5.2.9** Assume that \( f : K \to K' \) is a morphism. Then there exists a unique isomorphism \( f^* : K/\ker(f) \to K' \) with \( f = f^* \circ \eta_{\ker(f)} \).
5.3 Bisimulations

Proof Define \( f^\sharp := f_{\ker(f)} \), then the maps constituting this morphism are bijective Borel maps, so by [88, Proposition 4.5.1] they are Borel isomorphisms. The equations establishing the morphism property for \( f_{\ker(f)} \) show that the inverses also constitute a morphism. \( \dashv \)

Corollary 5.2.10 Let \( c \) and \( d \) be congruences on \( K \), then the following statements are equivalent:

1. \( c \preceq d \)
2. \( d = e \cdot c \) for some congruence \( e \) on \( K \).

Proof The implication 2 \( \Rightarrow \) 1 is obvious. Assume that \( c \preceq d = \ker(\eta_d) \) holds. Then the assertion follows from Proposition 5.2.8 together with Corollary 5.2.2. \( \dashv \)

This property is somewhat surprising in that it relates the refinement of congruences to factor spaces. If congruence \( c \) is finer than congruence \( d \), then \( d \) can be obtained through observing and factoring the behavior in the factor system for \( c \) (so that not the original system has to be observed but rather a simplified one).

5.3 Bisimulations

Bisimulations are introduced as spans of morphisms such that common events exist. They relate two systems in terms of their elements, hence in terms of non-deterministic relations of their state spaces. In fact, let \((S, (\rightarrow_a)_{a \in A})\) and \((S', (\rightarrow'_a)_{a \in A})\) be two labelled transition systems, then a relation \( R \subseteq S \times S' \) is called a bisimulation iff

- Whenever \( \langle s, s' \rangle \in R \) and \( s \rightarrow_a s_1 \), then there exists \( s'_1 \) with \( s' \rightarrow'_a s'_1 \) and \( \langle s_1, s'_1 \rangle \in R \).
- Whenever \( \langle s, s' \rangle \in R \) and \( s' \rightarrow'_a s'_1 \), then there exists \( s_1 \) with \( s \rightarrow_a s_1 \) and \( \langle s_1, s'_1 \rangle \in R \).

Interpreting a labelled transition system as a coalgebra \((S, \alpha_S)\) for the functor \( \mathfrak{F} := \mathfrak{F}_{\text{om}}(A \times \cdot) \). Rutten [77, Example 2.1] shows that \( R \) is a congruence iff there exists a coalgebraic structure \( \alpha_R \) on \( R \) such that this diagram commutes:

\[
\begin{array}{ccc}
S & \xrightarrow{\pi_S} & R & \xrightarrow{\pi_{S'}} & S' \\
\alpha_S & & \alpha_R & & \alpha_{S'} \\
\mathfrak{F}(S) & \xrightarrow{\mathfrak{F}(\pi_S)} & \mathfrak{F}(R) & \xrightarrow{\mathfrak{F}(\pi_{S'})} & \mathfrak{F}(S')
\end{array}
\]

In section 5.4 we will specialize the discussion to the case that the morphisms are projections, and relate the different notions of bisimulations to each other. The present section is devoted to the general case, which turns out to be rich enough.

Definition 5.3.1 The stochastic relations \( K = (X, Y, K) \) and \( L = (V, W, L) \) are called bisimilar iff there exists a stochastic relation \( M = (A, B, M) \), morphisms \( f = (\phi, \psi) : M \rightarrow L \), and \( g = (\gamma, \delta) : M \rightarrow L \) such that
1. the diagram

\[
\begin{array}{c}
X \quad \phi \quad A \quad \gamma \quad V' \\
\downarrow \quad \quad \quad \quad \quad \downarrow \\
\mathcal{S}(Y) \quad \mathcal{S}(\psi) \quad \mathcal{S}(B) \quad \mathcal{S}(\delta) \quad \mathcal{S}(W) \\
K \quad M \quad L
\end{array}
\]

is commutative,

2. the \(\sigma\)-algebra \(\psi^{-1}[\mathcal{B}(Y)] \cap \delta^{-1}[\mathcal{B}(W)]\) is non-trivial, i.e., contains not only \(\emptyset\) and \(\mathcal{B}\).

The relation \(M\) is called mediating.

The first condition on bisimilarity states that \(f\) and \(g\) form a span of Stoch-morphisms

\[
K \quad f \quad M \quad g \quad L,
\]

thus we have for each \(a \in A, D \in \mathcal{B}(Y), E \in \mathcal{B}(W)\) the equalities

\[
K(\phi(a))(D) = (\mathcal{S}(\phi) \circ M)(a)(D) = M(a)(\phi^{-1}[D])
\]

and

\[
L(\psi(a))(E) = (\mathcal{S}(\psi) \circ M)(a)(E) = M(a)(\psi^{-1}[E]).
\]

The second condition states that we can find an event \(C^* \in \mathcal{B}(B)\) which is common to both \(K\) and \(L\) in the sense that

\[
\psi^{-1}[D] = C^* = \delta^{-1}[E]
\]

for some \(D \in \mathcal{B}(Y)\) and \(E \in \mathcal{B}(W)\) such that both \(C^* \neq \emptyset\) and \(C^* \neq \mathcal{B}\) hold (note that for \(C^* = \emptyset\) or \(C^* = \mathcal{B}\) we can always take the empty and the full set, resp.). Given such a \(C^*\) with \(D\) and \(E\) from above we get for each \(a \in A\)

\[
K(\phi(a))(D) = M(a)(\psi^{-1}[D])
\]

\[
= M(a)(C^*)
\]

\[
= M(a)(\delta^{-1}[E])
\]

\[
= L(\gamma(a))(E),
\]

thus the event \(C^*\) ties \(K\) and \(L\) together. Loosely speaking, \(\psi^{-1}[\mathcal{B}(Y)] \cap \delta^{-1}[\mathcal{B}(W)]\) can be described as the \(\sigma\)-algebra of common events, which is required to be non-trivial. If \(Y = W\), another interpretation of common events is discussed in Lemma 5.1.12 in terms of invariant Borel sets, but the discussion there addresses different issues than in the present context.

Note that without the second condition two relations \(K\) and \(L\) which are strictly probabilistic (i.e., for which the entire space is always be assigned probability one) would always be bisimilar: Put \(A := X \times V, B := Y \times W\) and set for \(\langle x, v \rangle \in A\) as the mediating relation \(M(x, v) := K(x) \otimes L(v)\), then the projections will make the diagram commutative. It is also clear that this argument does not work for the sub-probabilistic case. This
5.3 Bisimulations

Curious behavior of probabilistic relations is a bit surprising, but these relations step out of line in other situations as well: e.g. it will be shown that the full subcategory of probabilistic relations in anStoch has a final object, while anStoch itself does not have one, see Section 5.5, in particular Corollary 5.5.7 and the discussion leading to it. The second condition in Definition 5.3.1 serves to prevent this somewhat anomalous behavior; it is technically not too restrictive, as we will see below.

An important instance of congruences and factor spaces is furnished through equivalent congruences.

**Definition 5.3.2** Let \( K = (X, Y, K) \) and \( K' = (X', Y', K') \) be Polish objects with congruences \( c = (\alpha, \beta) \) and \( c' = (\alpha', \beta') \), respectively.

1. Call \( c \) proportional to \( c' \) (symbolically \( c \propto c' \)) iff \( \alpha \) spawns \( \alpha' \) via \( (\Upsilon, A_0) \), \( \beta \) spawns \( \beta' \) via \( (\Theta, B_0) \) such that

\[
\forall x \in X \forall x' \in \Upsilon([x]_c) \forall B \in B_0 : K(x)(B) = K'(x')(\Theta_B).
\]

2. Call these congruences equivalent iff both \( c \propto c' \) and \( c' \propto c \) hold.

Thus equivalent congruences behave in exactly the same way. The same behavior is exhibited on each atom, i.e., equivalence class, as far as the input is concerned, and on the respective invariant output sets. It becomes visible now that a characterization of equivalent behavior through congruences betrays the double face of congruences: it is certainly necessary to use the equivalence relation on the input spaces; but since the behavior on the output spaces is modelled through probabilities, we need also the invariant Borel sets for a characterization.

We will show now how equivalent congruences on stochastic relations give rise to a factor object built on their sum. This construction will be of use in Proposition 5.3.3 for investigating the bisimilarity of stochastic relations.

Assume that \( c \) and \( c' \) are equivalent congruences on the Polish objects \( K = (X, Y, K) \), and \( K' = (X', Y', K') \), respectively. Construct for \( K \) and \( K' \) the direct sum

\[
K \oplus K' := (X + X', Y + Y', K \oplus K'),
\]

where the only non-obvious construction is \( K \oplus K' \): put for the Borel set \( E \subseteq Y + Y' \)

\[
(K \oplus K')(z)(E) := \begin{cases} 
K(z)(E \cap Y), & \text{if } z \in X \\
K'(z)(E \cap Y'), & \text{if } z \in X',
\end{cases}
\]

then clearly \( K \oplus K' : X + X' \sim Y + Y' \). Define on \( X + X' \) resp. \( Y + Y' \) the \( \sigma \)-algebras

\[
\mathcal{G} := \{ C + C' | C \in \mathcal{I}NV (\mathcal{B}(X), \alpha), C' \in \mathcal{I}NV (\mathcal{B}(X'), \alpha') \}
\]

\[
\mathcal{H} := \{ D + D' | D \in \mathcal{I}NV (\mathcal{B}(Y), \beta), D' \in \mathcal{I}NV (\mathcal{B}(Y'), \beta') \},
\]

then \( \mathcal{G} \) and \( \mathcal{H} \) are countably generated sub-\( \sigma \)-algebras of the respective Borel sets. Because the \( \sigma \)-algebras in question are countably generated, so is their sum, and because the congruences are equivalent, we claim that \( z (\alpha + \alpha') z' \) implies that \( (K \oplus K')(z)(F) = (K \oplus K')(z')(F) \) holds for all \( F \in \mathcal{H} \). To establish this, fix \( z \in X, z' \in X' \), and consider

\[
S := \{ F \in \mathcal{H} | (K \oplus K')(z)(F) = (K \oplus K')(z')(F) \}.
\]
Since the congruences are equivalent, this is a $\sigma$-algebra containing the generator $\{D_n + \Theta D_n | n \in \mathbb{N}\}$, where $\beta$ spawns $\beta'$ via $(\Theta, \{D_n | n \in \mathbb{N}\})$. Since the generator is closed under finite intersections, measures are uniquely determined by the $\pi$-$\lambda$-Theorem A.1.1. This implies $\mathcal{H} \subseteq \sigma(S)$, thus $\mathcal{H} = S$. Consequently,

$$
\mathcal{G} = \text{INV} (B(X + X'), \alpha + \alpha')
$$

$$
\mathcal{H} = \text{INV} (B(Y + Y'), \beta + \beta'),
$$

and $c + c' := (\alpha + \alpha', \beta + \beta')$ is a congruence on $K \oplus K'$.

The factor object $(K \oplus K')/(c + c')$ constructed in this way will be investigated more closely in Proposition 5.3.3 below. There we will establish that $K$ and $K'$ are bisimilar, provided they have equivalent non-trivial congruences.

Equivalent congruences give rise to bisimilar stochastic relations, thus if we are presented with two stochastic relations for which we can establish the existence of non-trivial congruences which are equivalent, then the relations are bisimilar. This is a rather far-reaching generalization of the by now well-known characterization of bisimilarity of labelled Markov transition systems through mutually equivalent states, which will be discussed at length in section 6.1. Note that we give here is an intrinsic characterization of bisimilarity: we investigate the relations and their congruences on their own, but we do not need an external instance (like a logic) to determine bisimilarity. The technical tool for establishing this property is the existence of semi-pullbacks, which we have established in Theorem 4.3.5.

**Proposition 5.3.3** If there exists non-trivial congruences $c_i$ on the Polish objects $K_i$ for $i = 1, 2$ that are equivalent, then $K_1$ and $K_2$ are bisimilar.

**Proof** 1. Assume $K_i = (X_i, Y_i, K_i)$ and $c_i = (\alpha_i, \beta_i)$ for $i = 1, 2$. Construct the sum $K_1 \oplus K_2$ as above, and let $(\kappa_i, \lambda_i)$ be the corresponding injections, which are, however, no morphisms. Let $(\eta_{\alpha_1 + \alpha_2}, \eta_{\beta_1 + \beta_2}) : K_1 \oplus K_2 \rightarrow (K_1 \oplus K_2)/(c_1 + c_2)$ be the canonical injection, then $(\eta_{\alpha_1 + \alpha_2} \circ \kappa_1, \eta_{\beta_1 + \beta_2} \circ \lambda_1)$ constitutes a morphism $K_i \rightarrow (K_1 \oplus K_2)/(c_1 + c_2)$, as will be shown now. Surjectivity has to be established, and we have to show that the $\sigma$-algebra of common events is non-trivial.

2. Each equivalence class $a \in (X_1 + X_2)/(\alpha_1 + \alpha_2)$ can be represented as $a = [x_1]_{\alpha_1} + [x_2]_{\alpha_2}$ for some suitably chosen $x_1 \in X_1, x_2 \in X_2$. Similarly, each equivalence class $b \in (Y_1 + Y_2)/(\beta_1 + \beta_2)$ can be written as $b = [y_1]_{\beta_1} + [y_2]_{\beta_2}$ for some $y_1 \in Y_1, y_2 \in Y_2$. Conversely, the sum of classes is a class again. This follows from Lemma 5.1.23.

3. Now we have the following diagram:

$$
\begin{array}{c}
K_1 \\
\downarrow \\
(K_1 \oplus K_2)/(c_1 + c_2)
\end{array}
$$

The semi-pullback of the pair of morphisms with a joint target constructed in the first step exists by Corollary 4.3.4, it is a Polish object $(A, B, M)$, where

$$
B = \{ (y_1, y_2) \in Y_1 \times Y_2 \mid [y_1]_{\beta_1 + \beta_2} = [y_2]_{\beta_1 + \beta_2} \}.
$$
We need to establish that there are indeed non-trivial common events. Since \(c\) is non-trivial, we can find an invariant Borel set \(D \in \mathcal{I}NV(B(Y_1), \beta_1)\) with \(\emptyset \neq D \neq Y_1\). Assume that \(\beta_1\) spawns \(\beta_2\) via \((\Theta, \{D_n \mid n \in \mathbb{N}\}\), then \(\emptyset \neq \Theta D \neq Y_2\) also holds. Because \(D\) is \(\beta_1\)-invariant,

\[
\pi_{1, Y_1}^{-1}[D] = \{(y_1, y_2) \mid y_1 \in D\} = \{(y_1, y_2) \mid y_2 \in \Theta D\} = \pi_{2, Y_2}^{-1}[\Theta D]
\]

thus

\[
\pi_{1, Y_1}^{-1}[D] \in \pi_{1, Y_1}^{-1}[\mathcal{B}(Y_1)] \cap \pi_{2, Y_2}^{-1}[\mathcal{B}(Y_2)],
\]

and we are done once it is shown that \(\pi_{1, Y_1}^{-1}[D] \neq B\). Since \(D \neq Y_1\) is invariant, there exists \(y_1\) with

\[
[y_1]_{\beta_1 + \beta_2} \cap D = [y_1]_{\beta_1} \cap D = \emptyset.
\]

Let \([y_2]_{\beta_2} := \Theta([y_1]_{\beta_1})\), then \([y_2]_{\beta_1 + \beta_2} \cap \Theta D = [y_2]_{\beta_2} \cap \Theta D = \emptyset\). Consequently, \((y_1, y_2) \in B \setminus \pi_{1, Y_1}^{-1}[D]\). This shows that \(\pi_{1, Y_1}^{-1}[\mathcal{B}(Y_1)] \cap \pi_{2, Y_2}^{-1}[\mathcal{B}(Y_2)]\) is non-trivial. \(\dagger\)

The proof's strategy is to make sure that the classes associated with the congruences are distributed evenly among the summands in the sense that each class in the sum is the sum of appropriate classes. This then implies that we can construct surjective maps, and from them morphisms through some general mechanisms. The idea works in particular with isomorphic factor spaces.

**Proposition 5.3.4** Let \(K\) and \(K'\) be analytic objects such that \(K/c\) is isomorphic to \(K'/c'\) for some non-trivial congruences \(c\) and \(c'\). Then \(K\) and \(K'\) are bisimilar.

**Proof 0.** Let \(K = (X, Y, K)\) with \(c = (\alpha, \beta)\), similar for \(K'\) and \(c'\). Assume that \(f = (\Phi, \Psi)\) is the isomorphism \(K/c \rightarrow K'/c'\) which is composed of the Borel isomorphisms \(\Phi : X/\alpha \rightarrow X'/\alpha'\) and \(\Psi : Y/\beta \rightarrow Y'/\beta'\). Let moreover \(A\) and \(B\) be countable generators of \(\mathcal{I}NV(B(X), \alpha)\) and \(\mathcal{I}NV(B(Y), \beta)\) which are closed under finite intersections. We know from Proposition 5.1.24 that \(\alpha\) spawns \(\alpha'\) via \((\Phi, \Lambda)\), and that \(\beta\) spawns \(\beta'\) via \((\Psi, \Lambda)\).

Hence we have to establish for each \(x \in X, x' \in \Phi([x]_\alpha)\) and for each \(\beta\)-invariant Borel subset \(B \subseteq Y\) that \(K(x)(B) = K'(x')(\Psi_B)\) holds. This will imply \(c \propto c'\), interchanging the rôles of \(c\) and \(c'\) then will yield the result.

1. Given \(B \in \mathcal{I}NV(B(Y), \beta)\) we know from Lemma 5.1.8 that we can find a Borel set \(B_1 \in \mathcal{B}(Y/\beta)\) such that \(B = \eta_{\beta}^{-1}[B_1]\). Since \(\Psi\) is a Borel isomorphism, we find \(B_2 \in \mathcal{B}(Y'/\beta')\) with \(B_1 = \Psi^{-1}[B_2]\). A routine calculation shows that \(\Psi_B = \eta_{\beta'}^{-1}[B_2]\). Now assume that \(x \in X, x' \in \Phi([x]_\alpha)\), then the following chain of equations is obtained from the argumentation above, and from the assumption that \(f\) is an isomorphism.

\[
K(x)(B) = K(x)(\eta_{\beta}^{-1}[\Psi^{-1}[B_2]])
= K_{\alpha, \beta}([x]_\alpha)(\Psi^{-1}[B_2])
= K'_{\alpha', \beta'}(\Phi([x]_\alpha))(B_2)
= K'(x')(\eta_{\beta'}^{-1}[B_2])
= K'(x')(\Psi_B).
\]

This establishes the desired relation \(c \propto c'\) and completes the proof. \(\dagger\)

Thus isomorphic factor spaces make sure that the relations are bisimilar. These factor spaces arise e.g., when considering blocks that partition the target spaces of a relation.
into pairs of pieces that have the same size each. This idea is expressed in the following definition.

**Definition 5.3.5** Let $K = (X,Y,K)$ and $L = (X,Z,L)$ be analytic objects. Call $J = \{(B_i,C_i) \mid i \in I\}$ a block for $K,L$ iff

1. the index set $I \neq \emptyset$ is at most countable,
2. $\{B_i \mid i \in I\}$ and $\{C_i \mid i \in I\}$ are partitions of $Y$ resp. $Z$ into Borel sets,
3. $K(x)(B_i) = L(x)(C_i)$ holds for all $x \in X$ and all $i \in I$.

Thus a block $J$ cuts $Y$ and $Z$ into the same number of non-empty pieces, and corresponding pieces have the same probability for all $x \in X$. The existence of a block makes sure that $K$ and $L$ are bisimilar.

**Corollary 5.3.6** The analytic objects $K = (X,Y,K)$ and $L = (X,Z,L)$ are bisimilar, provided there exists a block of size at least two for them.

**Proof** 0. We will show that the existence of a block enables us to construct isomorphic factor spaces for suitable non-trivial congruences. This will then imply the assertion through Proposition 5.3.4.

1. Let $J = \{(B_i,C_i) \mid i \in I\}$ be the block which contains at least two elements. The partition $\{B_i \mid i \in I\}$ induces a smooth equivalence relation $\beta$ on $Y$ such that the equivalence classes are exactly the partition elements. Thus $[y]_\beta = B_i$ iff $y \in B_i$. The invariant Borel sets for $\beta$ are isomorphic to the power set of $I$,

$$\mathcal{INV}(B(Y),\beta) = \left\{ \bigcup_{i \in I_0} B_i \mid I_0 \subseteq I \right\}.$$

This is so because $\mathcal{INV}(B(Y),\beta) = \sigma(\{B_i \mid i \in I\})$, and because the $B_i$ form a partition of $Y$.

Put $c := (\Delta_X,\beta)$, then $c$ is a non-trivial congruence for $K$. Computing the factor relation $K_{\Delta_X,\beta}$, we see that

$$K_{\Delta_X,\beta}(x)(E) = \sum_{i \in I_0} K(x)(B_i),$$

provided $\eta_\beta^{-1}[E] = \bigcup_{i \in I_0} B_i$ holds for the Borel set $E \in B(Y/\beta)$, see Proposition 5.1.9. Similarly, define the smooth equivalence $\gamma$ on $Z$ through the partition $\{C_i \mid i \in I\}$, then $d := (\Delta_X,\gamma)$ is a non-trivial congruence for $L$. We have

$$L_{\Delta_X,\gamma}(x)(F) = \sum_{i \in I_0} K(x)(C_i),$$

whenever $\eta_\gamma^{-1}[F] = \bigcup_{i \in I_0} C_i$ holds for the Borel set $F \subseteq Z/\gamma$.

2. Now team up each element in the partition for $Y$ with its partner in $Z$: put

$$\psi : Y/\beta \ni B_i \mapsto C_i \in Z/\gamma,$$

then $\psi : Y/\beta \to Z/\gamma$ is a measurable bijection, and it is immediate from the definition of a block that $(id_X,\psi) : K/c \to L/d$ is an isomorphism.\$
5.3 Bisimulations

**Example 5.3.7** Let $K : S \rightsquigarrow S$ and $L : S \rightsquigarrow S$ be stochastic relations over the analytic space $S$ such that

1. $K(s)(S) = L(s)(S) \neq 0$ for all $s \in S$,

2. there exists points $s_K \neq s_L$ in $S$ with $K(s)\{s_K\} = L(s)\{s_L\}$ for all $s \in S$.

Then $(S, S, K)$ and $(S, S, L)$ are bisimilar. This follows at once from Corollary 5.3.6, because

$$\{(s_K, s_L), (S \setminus s_K, S \setminus s_L)\}$$

is a block for these relations. ♦

We have seen in Proposition 5.3.4 that isomorphic factor spaces make sure that the relations are bisimilar. The natural question is whether or not the converse also holds: given bisimilar relations, do they have isomorphic factor spaces? A first step towards an answer is done in

**Proposition 5.3.8** If the Polish objects $K$ and $K'$ are bisimilar such that the mediating object is compact with continuous morphisms, then $K$ and $K'$ have isomorphic non-degenerate factor spaces.

**Proof** 1. Let

$$K \leftarrow^f M \longrightarrow^{f'} K'$$

be the span of morphisms constituting bisimilarity. Because $K$ is isomorphic to $M/\ker(f)$ by Corollary 5.2.9, we may restrict our attention to factors of $M$. Thus we assume that $K = M/c, K' = M/c'$, where both $c$ and $c'$ are the kernels of continuous morphisms. Suppose that we can find congruences $d$ and $d'$ such that $d \cdot c = d' \cdot c'$. Then

$$K/d = (M/c)/d$$

$$\cong M/d \cdot c \quad \text{(by Proposition 5.2.7)}$$

$$= M/d' \cdot c'$$

$$= (M/c')/d'$$

$$\cong K'/d',$$

($\cong$ indicating isomorphism) and we are done, provided $K/d$ is shown to be non-degenerate, or, equivalently, $d$ not to have the universal relation as its second component. When looking for suitable congruences $d$ and $d'$, in view of Corollary 5.2.10 it is sufficient to find a congruence $d' = (\gamma, \delta)$ with $c \leq d' \cdot c'$ for the given congruences $c$ and $c'$, and $\delta$ is not universal.

2. Assume $M = (X, Y, M)$, and suppose $c = (\alpha, \beta), c' = (\alpha', \beta')$. We know that there exist smooth equivalence relations $\gamma$ and $\delta$ with $\alpha \subseteq \gamma \cdot \alpha'$ and $\beta \subseteq \delta \cdot \beta'$, moreover we know for a $\delta$-invariant Borel subset $D \in INV(Y/\beta', \delta)$ that $\eta_{\beta'}^{-1}[D] \in INV(B(Y), \beta) \cap INV(B(Y), \beta')$. This was shown in Proposition 5.1.19 and Lemma 5.1.20.

We show that $d = (\gamma, \delta)$ is a congruence, thus we have to show that $K_{\alpha', \beta'}(s)(D) = K_{\alpha', \beta'}(s')(D)$, whenever $D$ is a $\delta$-invariant Borel subset of $Y/\beta'$, and $s \gamma s'$. Assume first that

$$\langle s, s' \rangle \in \gamma_0 := \{\langle t, t' \rangle \mid t, t' \in X/\alpha', t \times t' \cap \alpha \neq \emptyset\}.$$
Then we can find \( \langle x, x' \rangle \in \alpha \) such that \( s = [x]_{\alpha'}, s' = [x']_{\alpha'} \), and \( [x]_{\alpha} = [x']_{\alpha} \). Thus we obtain from \( D \)'s invariance properties

\[
K_{\alpha', \beta'}(s)(D) = K(x)(\eta_{\beta'}^{-1}(D)) = K(x')(\eta_{\beta'}^{-1}(D)) = K_{\alpha', \beta'}(s')(D).
\]

This means that the assertion is true for all \( \langle s, s' \rangle \in \gamma_0 \).

Now consider

\[
\hat{\gamma} := \{ \langle t, t' \rangle \mid t, t' \in X/\alpha', K_{\alpha', \beta'}(t)(D) = K_{\alpha', \beta'}(t')(D) \},
\]

then \( \hat{\gamma} \) is an equivalence relation which contains \( \gamma_0 \), and consequently it contains \( \gamma \), as the construction of \( \gamma \) as the transitive closure of \( \gamma_0 \) shows (see Section 5.1.3, Claim 4 on page 127).

3. Since \( M/c \) and \( M'/c' \) are bisimilar, we can find \( F \in I_{NV}(B(Y), \beta) \cap I_{NV}(B(Y), \beta') \) with \( \emptyset \neq F \neq Y \) (this is so since e.g. \( I_{NV}(B(Y), \beta) = \eta_{\beta}^{-1}[B(Y/\beta)] \) by Lemma 5.1.8). Now minimality of the construction leading to Proposition 5.1.19 enters the argumentation: from Lemma 5.1.20 we infer that

\[
\eta_{\beta'}^{-1}[I_{NV}(B(Y/\beta'), \delta)] = I_{NV}(B(Y), \beta) \cap I_{NV}(B(Y), \beta')
\]

holds, thus we can find \( F_0 \in I_{NV}(B(Y/\beta'), \delta) \) with \( \emptyset \neq F_0 \neq Y/\beta' \). Consequently, \( \delta \) is not universal, and we are done. \( \square \)

Summarizing, we have established the following characterization of bisimilarity through congruences:

**Theorem 5.3.9** Consider for analytic objects \( K \) and \( K' \) the statements

1. There exist non-trivial congruences \( c \) and \( c' \) on \( K \) resp. \( K' \) such that \( K/c \) and \( K'/c' \) are isomorphic.

2. \( K \) and \( K' \) are bisimilar.

Then 1 \( \Rightarrow \) 2 holds always, and 2 \( \Rightarrow \) 1 holds in case the mediating object is compact and the associated morphisms are continuous.

This is an intrinsic characterization of bisimilarity through congruences, because we can look at the stochastic relations and say that they are bisimilar without having to look at an external instance (like the sentences of a modal logic). It would be most valuable to lift the rather strong condition on compactness. The proofs given above, in particular in Section 5.1.3, Claims 1 through 4 rely on compact spaces via the possibility to extract a converging subsequence from each sequence (hence on sequential compactness, to be specific). Otherwise smoothness cannot be guaranteed, but smoothness is crucial since it makes sure that the factor space is analytic.

**Conjecture.** The characterization of bisimilarity through isomorphic factor spaces is valid for all stochastic relations over analytic spaces.
5.4 2-Bisimulations

A bisimulation between the stochastic relations $K_1$ and $K_2$ has been defined in Section 5.3 through a stochastic relation $M$ (the mediating object) together with two morphisms

$$
K_1 \xleftarrow{f_1} M \xrightarrow{f_2} K_2
$$

that share some common event, cp. Definition 5.3.1. If $K_1$ and $K_2$ coincide, this is called a bisimulation on $K_1$. Thus a bisimulation is a span of morphisms. We will in this Section specialize this notion: rather taking general morphisms, we consider projections. Consequently, we have domains and ranges of the mediating object as relations. These relations may have interesting properties, for example when we discuss bisimulations on a single relation. There the question arises whether or not the underlying sets are equivalence relations, are smooth etc.

Formally, let

$$M = (\langle A, \mathcal{X} \rangle, \langle B, \mathcal{Y} \rangle, M)$$

be the mediating object with suitable $\sigma$-algebras $\mathcal{X}$ and $\mathcal{Y}$ on $A$ resp. $B$. If $A$ and $B$ are measurable subsets of $X_1 \times X_2$ resp. $Y_1 \times Y_2$, and if $f_1 = (\pi_{1,X_1}, \pi_{1,Y_1}), f_2 = (\pi_{2,X_2}, \pi_{2,Y_2})$ — $\pi$ indicating the projections —, then the bisimulation is called a 2-bisimulation. Thus a 2-bisimulation renders this diagram commutative:

$$
\begin{array}{ccc}
X_1 & \xleftarrow{\pi_{1,X_1}} & A & \xrightarrow{\pi_{2,X_2}} & X_2 \\
K_1 \downarrow & & \downarrow M & & \downarrow K_2 \\
\mathcal{G}(Y_1, A_1) & \xleftarrow{\mathcal{G}(\pi_{1,Y_1})} & \mathcal{G}(B, \mathcal{Y}) & \xrightarrow{\mathcal{G}(\pi_{2,Y_2})} & \mathcal{G}(Y_2, B_2)
\end{array}
$$

We require for 2-bisimulations $A$ and $B$ only to be measurable subsets of $X_1 \times X_2$ resp. $Y_1 \times Y_2$, and the $\sigma$-algebras $\mathcal{X}$ and $\mathcal{Y}$ chosen in such a way that the projections are morphisms, i.e., surjective and measurable maps. Note also that the condition on a non-trivial $\sigma$-algebra of common events now reads that there exists Borel sets $C_1 \subseteq Y_1, C_2 \subseteq Y_2$ with

$$\emptyset \neq B \cap (C_1 \times Y_2) = B \cap (Y_1 \times C_2) \neq B.$$

Having congruences available permits specializing the notion of a bisimulation further (and these specializations will be used later on when characterizing simple systems).

**Definition 5.4.1** Let $\alpha$ and $\beta$ be smooth equivalence relations on $X$ resp. $Y$.

1. A 2-bisimulation $M = (\alpha, \beta, M)$ on $K$ is called a smooth 2-bisimulation on $K$.

2. If the stochastic relation $N = ((\alpha, B(\alpha)), (\beta, \otimes \mathcal{Y}, \beta), N)$ has the property that

$$
(\mathcal{G}(\pi_{1,Y}) \circ N(a_1, a_2))(E) = K(a_1)(E) \text{ and } (\mathcal{G}(\pi_{2,Y}) \circ N(a_1, a_2))(E) = K(a_2)(E)
$$

hold whenever $(a_1, a_2) \in \alpha$ and $E$ is a $\beta$-invariant Borel set of $Y$, then $N$ is called a weak 2-bisimulation on $K$. 

145
Let $\beta \neq U_Y$, then there exist a $\beta$-invariant Borel set $\emptyset \neq P \neq Y$ (since there exists $y_1, y_2 \in Y$ with $\langle y_1, y_2 \rangle \notin \beta$, one may take $P := [y_1]_\beta$). Because by Lemma 5.1.13, part 1,
$$\emptyset \neq \beta \cap (P \times Y) = \beta \cap (P \times P) = \beta \cap (Y \times P) \neq \beta,$$
we see that the $\sigma$-algebra of common events is in this case not empty.

**Smooth** 2-bisimulations correspond to the bisimulation equivalences studied in coalgebras [77], as we will see soon. **Weak** 2-bisimulations restrict their attention to the $\beta$-invariant Borel sets of $Y$ (rather than on all Borel sets), $N(a) ((B \times Y) \cap \beta)$ is defined for $a \in \alpha$ and for the Borel set $B \in \mathcal{INV} (B(Y), \beta)$, see Lemma 5.1.13, part 2. This looks of course much more restrictive than for a smooth 2-bisimulation: Clearly a smooth 2-bisimulation is a weak one, and we will show in Proposition 5.4.2 that we can even produce a smooth 2-bisimulation from a weak one, provided the relation $K$ is a Polish object.

We will begin with an observation relating congruences, smooth and weak 2-bisimulations.

**Proposition 5.4.2** Consider the following conditions:

a. $c = (\alpha, \beta)$ is a non-trivial congruence on $K$.

b. There exists $N : (\alpha, B(\alpha)) \rightsquigarrow (\beta, \otimes [Y, \beta])$ such that $((\alpha, B(\alpha)), (\beta, \otimes [Y, \beta]), N)$ is a weak 2-bisimulation on $K$.

c. There exists $M : \alpha \rightsquigarrow \beta$ such that $(\alpha, \beta, M)$ is a smooth 2-bisimulation on $K$.

Then the following holds:

1. $c \Rightarrow b \Rightarrow a$ is true for the analytic spaces $X$ and $Y$,

2. If both $X$ and $Y$ are Polish, then all conditions are equivalent.

**Proof** 0. $c \Rightarrow b$ is quite obvious, since each smooth 2-bisimulation is a weak one, so for the general case the implication $b \Rightarrow a$, and for the Polish case the implication $a \Rightarrow c$ needs to be established.

1. $b \Rightarrow a$: Let $C \in \mathcal{INV} (B(Y), \beta)$ be a $\beta$-invariant Borel subset of $Y$, then

$$(C \times Y) \cap \beta = (Y \times C) \cap \beta = (C \times C) \cap \beta$$

has been established in Lemma 5.1.13, part 1. Thus we obtain for $\langle x, x' \rangle \in \alpha$ the following chain of equations from $((\alpha, B(\alpha)), (\beta, \otimes [Y, \beta]), N)$ being a 2-bisimulation

$$K(x)(C) = K(\pi_{1,X}(x, x'))(C) = \mathcal{G} (\pi_{1,Y})(N(x, x'))(C) = N(x, x')( (C \times Y) \cap \beta) = N(x, x')( (Y \times C) \cap \beta) = \mathcal{G} (\pi_{2,Y})(N(x, x'))(C) = K(\pi_{2,X}(x, x'))(C) = K(x')(C).$$
2. \( a \Rightarrow c \): This part is harder. We need to construct a stochastic relation \( M : \alpha \sim \beta \) so that \((\alpha, \beta, M)\) forms a 2-bisimulation. The plan is very similar to the plan pursued for the existence of semi-pullbacks in Section 4.3.1, in particular for the proof of the central Lemma 4.3.1. Because there are subtle differences in the respective scenarios, we adapt the proof *mutatis mutandis*; the central arguments, however, remain in each case the same. The plan goes as follows: we show that this problem can be considered a selection problem as well. For this, we define on \( \alpha \) a suitable set-valued map \( \Gamma \) that takes on closed sets of measures on \( \beta \) and that satisfies the conditions of Proposition A.2.7 for the existence of a selector. The main difficulty will again lie in showing that \( \Gamma \) takes in fact non-empty values, and here invariant sets come in. Before doing all that, it is shown that the stage we are working on can be set up through closed sets and continuous maps. Since \( \beta \) is smooth, there exists a Polish space \( W \) and a Borel measurable map \( g : Y \to W \) such that \( \beta = \ker (g) \) by Lemma 5.1.6. We can find by Proposition A.2.1 a finer Polish topology on \( Y \) with the same the Borel sets \( B(Y) \) that makes \( g \) continuous. Thus \( \beta \) may be assumed to be a closed subset of \( Y \times Y \). Since \( Y \) is a Polish space, the space \( \mathcal{G}(Y) \) is Polish as well. Because \( \alpha \) is smooth, we find a Polish space \( V \) and a Borel measurable map \( h : X \to V \) such that \( \alpha = \ker (h) \). Applying the same argument as above, we can find a Polish topology on \( X \) which makes \( h : X \to Y \) as well as \( K : X \to \mathcal{G}(Y) \) continuous maps, rendering in particular \( \alpha \) a closed, hence Polish, subset of \( X \times X \).

Given \( \langle x_1, x_2 \rangle \in \alpha \), the set

\[
\Gamma(x_1, x_2) := \{ \mu \in \mathcal{G}(\beta) \mid \mathcal{G}(\pi_{1,Y})(\mu) = K(x_1), \mathcal{G}(\pi_{2,Y})(\mu) = K(x_2) \}
\]

will be scrutinized with the goal of finding a measurable selector for \( \Gamma \). It is immediate that it is a closed subset of \( \mathcal{G}(\beta) \), because the projections induce continuous maps on the respective spaces of sub-probabilities. Whenever \( C \subseteq \mathcal{G}(\beta) \) is compact, the weak inverse

\[
\exists \Gamma(C) := \{ \langle x_1, x_2 \rangle \in \alpha \mid \Gamma(x_1, x_2) \cap C \neq \emptyset \}
\]

of \( C \) is a closed subset of \( \alpha \): let \( \langle x_{1,n}, x_{2,n} \rangle_{n \in \mathbb{N}} \subseteq \exists \Gamma(C) \) be a sequence with

\[
\lim_{n \to \infty} \langle x_{1,n}, x_{2,n} \rangle = \langle x_1, x_2 \rangle.
\]

For each \( n \in \mathbb{N} \) there exists \( \gamma_n \in C \) with \( \gamma_n \in \Gamma(x_{1,n}, x_{2,n}) \). Since \( C \) is compact, there exists a subsequence \( s \) and \( \gamma \in C \) with \( \gamma_{s(n)} \to \gamma \). Continuity of \( K \) and closedness of \( \alpha \) together imply that \( \gamma \in \Gamma(x_1, x_2), \) thus \( \langle x_1, x_2 \rangle \in \exists \Gamma(C) \).

We show first that \( \Gamma(x_1, x_2) \neq \emptyset \), whenever \( \langle x_1, x_2 \rangle \in \alpha \). For this the techniques developed in Section 4.1 are used. Put \( Z := Y/\beta \) with \( C := B(Y/\beta) \), then \( (Z, C) \) is an analytic space, hence it is separable, see A.2.1. The map \( \psi : y \mapsto [y]_\beta \) is measurable from \( Y \) onto \( Z \), and we have

\[
S := \{ \langle y_1, y_2 \rangle \mid \psi(y_1) = \psi(y_2) \} = \beta.
\]

We know moreover from Proposition 5.1.9 that \( \eta^{-1}_\beta \{ C \} = I_{NV}(B(Y), \beta) \) holds. Now fix \( \langle x_1, x_2 \rangle \in \alpha \) and put \( \nu_1 := K(x_1), \nu_2 := K(x_2) \), then a measure \( \theta_1 \) on the \( \sigma \)-algebra \( \otimes [Y, \beta] \) is defined through

\[
(*) \theta_1((B \times B) \cap \beta) = \nu_1(B) (= \nu_2(B)),
\]

see Lemma 5.1.13, part 3. An appeal to Proposition 4.2.4 yields an extension of \( \theta_1 \) to a measure \( \theta \) which is defined on all of \( B(S) \). Thus we have now \( \theta \in \mathcal{G}(S) \) such that

\[
\forall E_i \in \psi^{-1} \{ C \} : \mathcal{G}((\pi_{i,Y})(\theta))(E_i) = \nu_i(E_i), i = 1, 2.
\]
From Proposition 4.2.4 we obtain a measure \( \mu \in \mathcal{S}(S) \) such that
\[
\forall E_i \in \mathcal{B}(S) : \mathcal{S}(\pi_{i,Y})(\mu)(E_i) = \nu_i(E_i), \quad i = 1, 2.
\]
But this means that \( \Gamma(x_1, x_2) \neq \emptyset \), thus we can apply the Selection Theorem (Proposition A.2.7) and obtain a measurable selector \( M \) for \( \Gamma \), consequently, \( M : \alpha \to \beta \). Thus \( M := (\alpha, \beta, M) \) is a stochastic relation. From \( M \) being a selector to \( \Gamma \) one sees that \( M \) is a 2-bisimulation for \( K \), since
\[
(\mathcal{S}(\pi_{1,Y}) \circ M)(x_1, x_2) = K(x_1)
\]
\[
(\mathcal{S}(\pi_{2,Y}) \circ M)(x_1, x_2) = K(x_2)
\]
is true for all \( (x_1, x_2) \in \alpha \).

Thus we have established a very close relationship between congruences and 2-bisimulations for stochastic relations. The basic idea has been again to extend a stochastic relation that is defined on a small and fairly easy to handle \( \sigma \)-algebra to a larger one. But this is complicated, because we do not have direct access to the Borel sets, when we need it: the Borel sets are defined in terms of a closure operation and not through some explicit procedure, so we cannot put a handle on them directly (in fact, this is a white lie: the Borel sets can be defined stepwise through transfinite induction, see e.g. [88] or [6]; but this process is rather complicated and will not help us here at all). Hence we have to walk a by-path again: we show through a selection argument that such a measure must exist.

Albeit there are subtle variations here and in Section 4.3.1, this argument works essentially as follows:

1. We know that the situation is easily managed on a the small \( \sigma \)-algebra which we start from (this is like the begin of a proof by induction: the picture is nice and clear in the beginning).

2. We know also that our request for an extension is not unreasonable, since our map \( \Gamma \) has some reasonable properties (this is like the induction hypothesis).

3. From this we conclude that we can find an extension through a selector (this is much like the inductive step itself).

Quite apart from the involved technical development, this close relationship between bisimilarity and congruences is somewhat akin to the scenario for general coalgebras. The situation cannot be mirrored, however, since for coalgebras one usually requires a functor which preserves weak pullbacks, see e.g. [77]. The structure for the sub-probability functor \( \mathcal{S} \) is slightly more involved, because the hope for establishing weak pullbacks is vain. Consequently it seems to be difficult to fit general coalgebras and stochastic relations too tightly under one common roof.

Anyway, Proposition 5.4.2 provides us with a considerable degree of freedom. It will be of use when investigating simple relations: we can select the proper scenario in investigating simplicity without having to be afraid that we lose important properties, as will be seen in Section 5.5. This holds at least in the Polish case. In the case of an analytic object we have to be a bit careful, but Proposition 5.4.2 tells us as well where to install watch dogs.

A partial converse to Proposition 5.4.2 is furnished through
Lemma 5.4.3 Let $\alpha$ and $\beta$ be smooth equivalence relations on $X$ resp. $Y$. Assume that

$$M := ((\alpha, B(\alpha)), (\beta, \otimes [Y, \beta]), M)$$

is a weak 2-bisimulation on $K$. Then $(\alpha, \beta)$ is a congruence of $K$.

Proof Let $B \in INV (B(Y), \ell(\beta))$, then we know from Lemma 5.1.13, part 1 that $(B \times Y) \cap \beta = (Y \times B) \cap \beta$ holds. Thus we get from the assumption that $M$ is a bisimulation on $K$ the equality $K(x_1)(B) = K(x_2)(B)$ for all $(x_1, x_2) \in \alpha$. ⊣

Proposition 5.4.2 builds the much needed bridge between congruences and bisimulations. Quite apart from being of considerable interest unto its own, we will cross this bridge when investigating simple systems.

5.5 Simple Relations

An algebraic structure which is isomorphic to each of its non-trivial factor spaces is called simple. Take e.g. a simple and non-trivial group $G$ and an epimorphism $\phi : G \to H$, then $\phi$ is an isomorphism, cp. [58, p. 104]. Since simple systems do not have non-trivial subsystems, a system $S$ is simple if each epimorphism $S \to T$ is an isomorphism. The very close connection between simple systems and trivial bisimulations is well-known in the theory of coalgebras: a system is simple iff it has only trivial bisimulations.

Simple systems will be characterized both for Polish and analytic spaces. We deal first with the Polish case which is a bit easier to handle, and then to the analytic case. A technique for reducing the analytic to the Polish case is developed, so that we may capitalize on previous results. A complete characterization of simple relations can be given for the analytic case.

Call a congruence $c = (\alpha, \beta)$ on $X$ and $Y$ plain iff both equivalence relations are the identity, viz., iff both $\alpha = \Delta_X$ and $\beta = \Delta_Y$ hold. Similarly, call a smooth or weak 2-bisimulation plain iff the underlying congruence is plain.

Definition 5.5.1 A stochastic relation $K$ is called simple iff each morphism with domain $K$ is an isomorphism.

This definition looks a bit stronger than usual, since usually epimorphisms emanating from a simple structure are assumed to be isomorphisms. But since all our morphisms are epis, we deal only with surjective maps, thus the common definition applies in this context as well.

Looking first at relations based on Polish spaces, things are rather satisfyingly characterized through equivalences of smooth and weak 2-bisimulations and plain congruences. There is even a characterization through morphisms going into the relation in question. The case of analytic relations is a bit more involved, since the equivalence of smooth and weak 2-bisimulations is not guaranteed, so it is relegated to a separate discussion.
5.5.1 The Polish Case

We characterize simple systems if both spaces on which the relation is defined are Polish. The following characterization is summarized in Figure 5.1.

**Theorem 5.5.2** Consider these statements for the Polish object $K$

(a). $K$ is simple.

(b). Each smooth 2-bisimulation on $K$ is plain.

(c). Each weak 2-bisimulation on $K$ is plain.

(d). Let $f_1, f_2: M \to K$ be morphisms, where $M$ is a Polish object, then $f_1 = f_2$.

(e). Each congruence on $K$ is plain.

Then

1. These implications hold always: $(a) \Leftrightarrow (b) \Leftrightarrow (c) \Leftrightarrow (e) \Leftrightarrow (d)$.

2. Let in (d) $f_i = (\phi_i, \psi_i)$. If both $\ell([\phi_1|\phi_2])$ and $\ell([\psi_1|\psi_2])$ are smooth, then $(e) \Rightarrow (d)$ holds as well.

The proof for Theorem 5.5.2 is broken into several pieces:

(c) $\Rightarrow$ (a): Let $f: K \to L$ be a morphism, then $f$ can be factored through $K/\ker(f)$ as $f = f' \circ \eta_{\ker(f)}$ with an isomorphism $f'$ by Corollary 5.2.9. $\ker(f)$ is a congruence which is plain by assumption. Thus $f$ is an isomorphism.

(a) $\Rightarrow$ (e): If $c$ is a congruence on $K$, then $\eta_c: K \to K/c$ is a morphism.

(b) $\leftrightarrow$ (e) This is a special case of Proposition 5.4.2.

(d) $\Rightarrow$ (b): Let $M := (A, B, M)$ be a smooth bisimulation on $K$, then

$$\pi_{1,X}, \pi_{1,Y}, (\pi_{2,X}, \pi_{2,Y}) : M \to K$$

are morphisms which are equal by assumption.

This settles the proof of part 1. Turning to the proof of part 2, assume that (e) holds in addition to $(\ell([\phi_1|\phi_2]), \ell([\psi_1|\psi_2]))$ being smooth. We note from the proof of Lemma 5.1.12 that a $\ell([\psi_1|\psi_2])$-invariant Borel set $D \subseteq Y$ has the property that $\psi_1^{-1}[D] = \psi_2^{-1}[D]$
holds, hence that $D$ is an event common to $\psi_1$ and $\psi_2$. Now define the equivalence relation

$$R_D := \{ (x_1, x_2) \mid K(x_1)(D) = K(x_2)(D) \},$$

then $[\psi_1||\psi_2] \subseteq R_D$ follows from $f_1, f_2 : M \to K$ being morphisms: suppose $\langle x_1, x_2 \rangle = \langle \phi_1(a), \phi_2(a) \rangle$, and $E = \psi_1^{-1}[D] = \psi_2^{-1}[D]$, we obtain

$$K(x_1)(D) = (K \circ \phi_1)(a)(D) = (\mathcal{G} (\psi_1) \circ M)(a)(D) = M(a)(E) = (\mathcal{G} (\psi_2) \circ M)(a)(D) = K(x_2)(D).$$

Since $R_D$ is an equivalence relation for each $D$, and $\ell([\phi_1||\phi_2])$ is the smallest equivalence relation containing $[\phi_1||\phi_2]$, this implies

$$\ell([\phi_1||\phi_2]) \subseteq \bigcap \{R_D \mid D \in INV (B(Y), \ell(T))\}$$

which in turn yields that $(\ell([\phi_1||\phi_2]), \ell([\psi_1||\psi_2]))$ is a congruence on $K$. This congruence is plain by assumption, yielding $f_1 = f_2$, as desired.

### 5.5.2 The Analytic Case

We will reduce the case of relations on analytic spaces to the one where we have Polish spaces at our disposal, and we have seen that we can move smooth equivalence relations along arrows (albeit reversing the direction) in Lemma 5.1.11. This will be used now to move congruences.

**Proposition 5.5.3** Let $K = (X, Y, K)$ be a Polish object, $L = (A, B, L)$ be an analytic object, assume that $f = (\phi, \psi) : K \to L$ is a morphism, and that $c = (\alpha, \beta)$ is a congruence on $L$. Then $c_f := (\alpha_\phi, \beta_\psi)$, is a congruence on $K$.

**Proof** We know from the constructions that both $\alpha_\phi$ and $\beta_\psi$ are smooth equivalence relations. Now let a $\alpha_\phi$ $a'\phi$, thus $\phi(a) \alpha \phi(a')$. Assume that $E \subseteq Y$ is a $\beta_\psi$-invariant Borel set; from Lemma 5.1.11 we may infer that $E = \psi^{-1}[E_0]$ for some $\beta$-invariant Borel set $E_0 \subseteq B$. Then

$$K(a)(E) = K(a)\psi^{-1}[E_0] = (\mathcal{G} (\psi) \circ K) (a)(E_0) = L(\phi(a))(E_0) = L(\phi(a'))(E_0) = K(a)(E),$$

because $(\phi, \psi)$ is a morphism, and because $(\alpha, \beta)$ is a congruence on $L$. This shows that $c_f$ is in fact a congruence on $K$. $\dashv$

For a characterization of simple stochastic relations analogous to Theorem 5.5.2, we fix an analytic object $K = (X, Y, K)$ together with Polish spaces and surjective Borel maps $f : X_0 \to X$ and $g : Y_0 \to Y$ which define the analytic structure on $X$ resp. $Y$. We establish for $K$ the following property:
**Proposition 5.5.4** These conditions are equivalent for \( K \):

1. Each weak 2-bisimulation on \( K \) is plain.

2. Each congruence on \( K \) is plain.

**Proof** 0. Since each weak 2-bisimulation is defined on a congruence, the implication 
\( 2 \Rightarrow 1 \) is obvious from Lemma 5.4.3. In order to establish the other implication, we will construct from a given congruence \( c = (\alpha, \beta) \) on \( K \) together with the derived pair \( c_{f,g} := (\alpha_f, \beta_g) \) a stochastic relation \( K_0 := (X_0, Y_0, K_0) \) on which \( c_{f,g} \) is a congruence, then construct a smooth 2-bisimulation \( M_0 = (\alpha_f, \beta_g, M_0) \) on \( K_0 \), and use this for constructing a weak 2-bisimulation \( M = (\alpha, \beta, M) \) on \( K \).

1. The relations \( \alpha_f \) and \( \beta_g \) are smooth equivalence relations on \( X_0 \) resp. \( Y_0 \). Define for \( E \in INV(B(Y), \beta) \) and \( x_0 \in X_0 \)

\[
K'_0(x_0)(g^{-1}[E]) := K(f(x_0))(E),
\]

then we see from Lemma 5.1.11 that \( K'_0 : (X_0, B(X_0)) \rightsquigarrow (Y_0, INV(B(Y_0), \beta_g)) \) is a stochastic relation, so by Proposition 4.2.7 we can find a stochastic relation

\[
K_0 : (X_0, B(X_0)) \rightsquigarrow (Y_0, B(Y_0))
\]

extending \( K'_0 \). Then \( c_{f,g} \) is a congruence on \( K_0 \); let \( \langle x_0, x_1 \rangle \in \alpha_f \), and \( E_0 \in INV(B(Y_0), \beta_g) \) be an invariant Borel set in \( Y_0 \). We know then that \( \langle f(x_0), f(x_1) \rangle \in \alpha \), and that \( E_0 = g^{-1}[g[E_0]] \) with \( g[E_0] \in INV(B(Y), \beta) \). Hence

\[
K_0(x_0)(E_0) = K_0(x_0)(g^{-1}[g[E_0]]) = K(f(x_0))(g[E_0]) = K(f(x_1))(g[E_0]) = K_0(x_1)(E_0).
\]

From Proposition 5.4.2 we get a smooth 2-bisimulation \( M_0 = (\alpha_f, \beta_g, M_0) \) on \( K_0 \). We show that this implies

\[
M_0(x_0, x_1)((P \times Y_0) \cap \beta_g) = M_0(x'_0, x'_1)((P \times Y_0) \cap \beta_g),
\]

provided \( P \in INV(B(Y_0), \beta_g) \) is a \( \beta_g \)-invariant Borel set in \( Y_0 \), and \( \langle x_0, x_1 \rangle, \langle x'_0, x'_1 \rangle \in \alpha_f \) with \( f(x_0) = f(x'_0) \) or \( f(x_1) = f(x'_1) \). This is done through the bisimulation property for \( M_0 \); assume that \( f(x_0) = f(x'_0) \), then

\[
M_0(x_0, x_1)((P \times Y_0) \cap \beta_g) = M_0(x_0, x_1)(\pi_{1,Y_0}^{-1}[P]) = (\mathcal{S}(\pi_{1,Y_0}) \circ M_0)(x_0, x_1)(P) = (K_0 \circ \pi_{1,Y_0})(x_0, x_1)(P) = K_0(x_0)(P) \overset{(*)}{=} K_0(x'_0)(P) = M_0(x'_0, x'_1)((P \times Y_0) \cap \beta_g).
\]

Eq. \( (*) \) follows from the observation that \( f(x_0) = f(x'_0) \) implies \( \langle x_0, x'_0 \rangle \in \alpha_f \) (note that \( x_1, x'_1 \) make sure that the respective arguments lie in the domain of \( M_0 \)).
Now introduce the stochastic relation $M = ((\alpha, B(\alpha)), (\beta, \otimes [Y, \beta]), M)$ by defining the sub-probability

$$M(a, a')(B \times Y) = M_0(x_0, x'_0)((g^{-1}[B] \times Y_0) \cap \beta_g)$$

for $\langle a, a' \rangle = \{f(x_0), f(x'_0)\} \in \alpha$ and for $B \in INV(B(Y), \beta)$.

The discussion above shows that $M$ is well defined, provided we can establish that $(g^{-1}[B] \times Y_0) \cap \beta_g \in \otimes [Y_0, \beta_g]$ is true. But we know that $g^{-1}[B] \in INV(B(Y_0), \beta_g)$ holds.

2. It remains to show that $M$ is indeed a weak 2-bisimulation. Let $\langle a, a' \rangle \in \alpha$ with $a = f(x_0), a' = f(x'_0)$, and take a $\beta$-invariant Borel set $E \subseteq Y$. Then $\pi_{1,Y}^{-1}[E] = (E \times Y) \cap \beta$, thus putting all this together, we obtain

$$M(a, a')((\pi_{1,Y}^{-1}[E])) = M(f(x_0), f(x'_0))((E \times Y) \cap \beta)$$
$$= M_0(x_0, x'_0)((g^{-1}[E] \times Y) \cap \beta_g)$$
$$= K_0(x_0)(g^{-1}[E])$$
$$= K(f(x))(E)$$
$$= K(a)(E).$$

As a consequence the analogue to Theorem 5.5.2 is obtained for analytic objects, see Figure 5.2 for a pictorial summary.

**Theorem 5.5.5** Consider these statements for the analytic object $K$

(a). $K$ is simple.

(b). Each smooth 2-bisimulation on $K$ is plain.

(c). Each weak 2-bisimulation on $K$ is plain.

(d). Let $f_1, f_2 : M \to K$ be morphisms, where $M$ is an analytic object, then $f_1 = f_2$.

(e). Each congruence on $K$ is plain.

Then these implications hold: $\text{(d)} \Rightarrow \text{(a)} \Leftrightarrow \text{(e)} \Rightarrow \text{(b)} \Rightarrow \text{(c)} \Leftrightarrow \text{(e)}$.

We are now in a position to characterize simple systems over analytic spaces completely. Let $\mathbb{1} := \{\ast\}$ be the one-element space with the discrete topology (which is Polish) and $\mathcal{P}(\mathbb{1})$ as its Borel sets. This space plays a distinguished rôle:
**Proposition 5.5.6** The analytic objects \((X, \mathbb{1}, K)\) such that \(x \mapsto K(x)(\mathbb{1})\) is injective are exactly the simple analytic objects.

**Proof** 1. Let \(K = (X, \mathbb{1}, K)\) be such an object, and assume that \(f = (\phi, \psi) : K \to L = (A, B, L)\) is a morphism. Then \(B\) can have only one element. Since \(x \mapsto K(x)(\mathbb{1})\) is one-to-one, we see that \(x \neq x'\) implies \(L(\phi(x))(B) \neq L(\phi(x'))(B)\), hence \(\phi(x) \neq \phi(x')\). Consequently \(f\) is an isomorphism. Thus \(K\) is simple.

2. Let conversely \(K = (X, Y, K)\) be a simple stochastic relation, and define the smooth equivalence relation \(\bar{\alpha}\) through \(\bar{\alpha} := \ker (K(\cdot)(Y))\). Put \(\bar{\omega} := Y \times Y\), then \(\bar{\omega}\) is also smooth. It is not difficult to see that \(c := (\bar{\alpha}, \bar{\omega})\) is a congruence on \(K\), since \(\{\emptyset, Y\}\) is the \(\sigma\)-algebra of \(\bar{\omega}\)-invariant subsets. By Theorem 5.5.5, \(c\) is plain, thus \(\bar{\alpha} = \Delta_X\) and \(\bar{\omega} = \Delta_Y\). Hence \(Y\) can only have one element, and \(\alpha\) is the kernel of an injective map. Thus the simple objects in the category of stochastic relations over analytic spaces are in one-to-one correspondence with the injective Borel maps from analytic spaces to the unit interval. Proposition 5.5.6 is the stochastic counterpart to the coalgebraic characterization of simple system which says that a system \(S\) is simple iff it is isomorphic to \(S/\equiv\), where \(\equiv\) is the greatest bisimulation on \(S\) [77, Theorem 8.1]. This construction is not directly applicable in the present context since the notion of a greatest bisimulation is not available here.

Call finally an object \(F\) final iff given another object \(M\) there exists exactly one morphism \(f : M \to F\). In view of Theorem 5.5.2, a final object is simple. The category of stochastic relations does not have final objects: Being simple, according to Proposition 5.5.6 a final object would have the shape \(F = (X, \mathbb{1}, F)\). But \(X\) cannot have more than one element, thus \(F = (\mathbb{1}, \mathbb{1}, F)\) with \(F(\mathbb{1}) = \{r\}\) for some \(r, 0 \leq r \leq 1\). But then there would be a unique morphism \((\mathbb{1}, \mathbb{1}, K) \to (\mathbb{1}, \mathbb{1}, K')\) with \(K'(\mathbb{1}) = r' \neq r\). This is evidently impossible.

We have, however, the following positive result:

**Corollary 5.5.7** The full subcategory of stochastic relations \((X, Y, K)\) such that \(K(x)(Y) = 1\) holds for all \(x \in X\) has a final object \((\mathbb{1}, \mathbb{1}, F)\).

### 5.6 Case Study: The Converse of a Stochastic Relation

Bisimilarity is quite robust a relation; this will be demonstrated for the converse of a stochastic relation. Quite apart from this observation, the problem is interesting in its own right, because it suggests an occasion for investigating some similarities between forming the converse for set-theoretic relations and their stochastic cousins. It is shown how the converse is constructed through a disintegration argument (in marked contrast to the set-theoretic case, where merely the order of the pairs needs to be reversed).

For introducing the problem, let \(R\) be a set-relation on a set of states. If \(\langle x, y \rangle \in R\), then this can be written as \(x \rightarrow_R y\) and interpreted as a state transition from \(x\) to \(y\). The converse \(R^\rightarrow\) shifts attention to the goal of the transition: \(y \rightarrow_{R^\rightarrow} x\) is interpreted as \(y\) being the goal of a transition from \(x\).

Now let \(p(x, y)\) be the probability that there is a transition from \(x\) to \(y\), and the question arises with which probability state \(y\) is the goal of a transition from \(x\). This question cannot be answered unless we know the initial probability \(\mu\) for the states. Then we can
5.6 Case Study: The Converse of a Stochastic Relation

Figure 5.3: A Stochastic Relation and Its Converse

calculate \( p_\mu \overset{\sim}{\cdot} (y, x) \) as the probability to make a transition from \( x \) to \( y \) weighted by the probability to start from \( x \) conditional to the event to reach \( y \) at all, i.e.

\[
p_\mu \overset{\sim}{\cdot} (y, x) := \frac{\mu(x) \cdot p(x, y)}{\sum_t \mu(t) \cdot p(t, y)}.
\]

Consider as an example the simple transition system \( p \) on three states given in the left hand side of Fig. 5.3. The converse \( p_\mu \overset{\sim}{\cdot} \) for the initial probability \( \mu := [1/2 \ 1/4 \ 1/4] \) is given on the right hand side.

The transition probabilities \( p \) are given through

\[
\begin{bmatrix}
1/4 & 1/2 & 1/4 \\
1/5 & 1/2 & 3/10 \\
1/3 & 1/3 & 1/3
\end{bmatrix}
\]

with initial probabilities according to the stochastic vector \( \mu := [1/2, 1/4, 1/4,] \). The converse \( p_\mu \overset{\sim}{\cdot} \) is then computed as

\[
\begin{bmatrix}
15/31 & 6/31 & 10/31 \\
6/11 & 3/11 & 2/11 \\
15/34 & 9/34 & 5/34
\end{bmatrix}.
\]

The situation is of course more complicated in the non-finite case. We assume as usual that we work in Polish spaces. A definition of the converse \( K_\mu \overset{\sim}{\cdot} \) of a stochastic relation \( K \) given an initial distribution \( \mu \) is proposed in terms of disintegration. An interpretation of the converse in terms of random variables is given, and it is shown that the converse behaves with respect to composition like its set-theoretic counterpart, viz., \( (K*L)_\mu \overset{\sim}{\cdot} = L_{(K^*\mu)}^*K_\mu \overset{\sim}{\cdot} \), where \( K^*\mu \) denotes as in Lemma 2.4.4 the image distribution of \( \mu \) under \( K \), and the composition is the Kleisli composition for the corresponding monad.
(section 5.6.1). This is of course the probabilistic counterpart to the corresponding law for relations $R$ and $S$, which reads $(R \ast S)^\sim = S^\sim \ast R^\sim$.

The set $\{K_\mu^\sim(y) \mid y \in Y\}$ of all sub-probability measures constituting the converse turns out to have an interesting property: it is topologically rather small, i.e., its closure is compact in the weak topology of sub-probability measures on $Y$, indicating that the converse $K_\mu^\sim$ does not carry as much information as $K$ or $\mu$ do.

Before entering the discussion on the converse, we study briefly the interplay between stochastic relations and the measures on the codomain. These operations will be helpful in the sequel, we look at them from a slightly different angle than in section 2.4.2.

**Lemma 5.6.1** Let $X$ and $Y$ be Polish spaces, $K : X \leadsto Y$ be a stochastic relation.

1. $K^*$ defines a map $K^* : \mathcal{S}(X) \rightarrow \mathcal{S}(Y)$ (see Lemma 2.4.4) such that
   \[
   \int_Y g \, dK^*(\mu) = \int_X \int_Y g(y) \, K(x)(dy) \, \mu(dx)
   \]
   holds for each $g \in \mathcal{F}(Y)$.

2. $(\mu \otimes K)(D) := \int_X K(x)(D_x) \, \mu(dx)$ assigns $\mu \in \mathcal{S}(X)$ and $K$ a sub-probability on $X \times Y$ such that
   \[
   \int_{X \times Y} g \, d(\mu \otimes K) = \int_Y \int_X g(x,y) \, K(x)(dy) \, \mu(dx)
   \]
   is true whenever $g \in \mathcal{F}(X \times Y)$.

**Proof** The proofs work along the following pattern, so often encountered here already: One first shows that the claim is correct for the case of indicator functions, then establishes that things work as expected for step functions as the linear combinations of indicator functions. Using a monotone approximation for non-negative bounded and measurable functions, the integral’s monotone continuity shows that the claim is justified for these functions; finally, a decomposition of a map into the difference of non-negative functions yields the claim for general measurable and bounded maps. The reader is invited to fill in the details.

Let us illustrate these constructions for the discrete case.

**Example 5.6.2** Assume that $p : \{1, \ldots, n\} \leadsto \{1, \ldots, m\}$ is a stochastic relation, and let $\mu \in \mathcal{S}(\{1, \ldots, n\})$ be an initial distribution. Then

1. $p^*(\mu)(j) = \sum_{i=1}^n \mu(i) \cdot p(i,j)$ is the probability that response $j$ is produced, given the initial probability $\mu$.

2. $(\mu \otimes p)((i,j)) = \mu(i) \cdot p(i,j)$ gives the probability for the input/output pair $(i, j)$ to occur, given the initial probability $\mu$ (which is responsible for input $i$), and the probability $p(i,j)$ for output $j$ after input $i$.

These properties are easily established using elementary computations.

It is remarkable that the construction in part 2 of Definition 5.6.1 can be reversed, and this is in fact the cornerstone for constructing the converse of a stochastic relation. **Reversing the construction** means that each measure on the product of two Polish spaces can be represented as a product of a stochastic relation with a measure.
Proposition 5.6.3 Given $\nu \in \mathcal{S}(X \times Y)$ there exists $\mu \in \mathcal{S}(X)$ and $K : X \rightsquigarrow Y$ with $\nu = \mu \otimes K$.

Proof This is but a reformulation of Proposition 4.2.6. ⊥

The stochastic relation $K$ is known as the regular conditional distribution of $\pi_Y$ given $\pi_X$, see section 4.2.2. Relation $K$ is sometimes called a version of the disintegration of $\zeta$ w.r.t. $\mathcal{S}(\pi_X \times Y, X)$ ($\zeta$).

Example 5.6.4 Let $\zeta \in \mathcal{S}(\{1, \ldots, n\} \times \{1, \ldots, m\})$, then the probability $p(i, j)$ for input $i$ generating output $j$ is the probability $\zeta((i, j))$ for the pair $(i, j)$ to occur conditioned on the probability $\sum_{t=1}^{m} \zeta((i, t))$ that input $i$ is produced at all. Thus relation $p$ satisfies the equation

$$\zeta((i, j)) = \left( \sum_{t=1}^{m} \zeta((i, t)) \right) \cdot p(i, j).$$

This is the discrete version of Proposition 4.2.6. In contrast to the discrete case, however, the version of the disintegration of $\zeta$ with respect to its projection usually cannot be computed explicitly in the general case. ◇

There is a rather helpful interplay between the projection of $\mu \otimes K$ to the second component and $K^*(\mu)$ which will be exploited later on.

Lemma 5.6.5 If $\mu \in \mathcal{S}(X)$ is a sub-probability measure, and $K : X \rightsquigarrow Y$ is a stochastic relation, then

$$\mathcal{S}(\pi_X \times Y, X) (\mu \otimes K) = K^*(\mu).$$

Proof Let $B \subseteq Y$ be a Borel set, then

$$\mathcal{S}(\pi_X \times Y, X) (\mu \otimes K)(B) = (\mu \otimes K)(X \times B) = \int_X K(x)((X \times B)_x) \mu(dx) = \int_X K(x)(B) \mu(dx) = K^*(\mu)(B).$$

5.6.1 Converse Relations

Given a sub-stochastic matrix $(p(i, j))_{1 \leq i \leq n, 1 \leq j \leq m}$ representing a stochastic relation $\{1, \ldots, n\} \rightsquigarrow \{1, \ldots, m\}$ and an initial distribution, we have seen above that the probability $p_{\mu}^{-}(j)(i)$ of responding with $j \in \{1, \ldots, m\}$ on a stimulus $i \in \{1, \ldots, n\}$ is calculated as

$$p_{\mu}^{-}(j)(i) = \frac{\mu(i) \cdot p(i, j)}{\sum_{t} \mu(t) \cdot p(t, j)}.$$

The probability $p_{\mu}^{-}$ under consideration reverses $p$ given an initial distribution, so is regarded as the converse of $p$ (inverse might at first sight be considered a better name, but this seems to suggest invertibility of the matrix associated with $p$).
In view of Examples 5.6.4 and 5.6.2, this amounts to the disintegration of \( \mu \otimes p \) with respect to the distribution \( p^*(\mu) = \mathcal{S}(\pi_{X \times Y})(\mu \otimes p) \).

This observation guides the way for the definition of the converse for a general stochastic relation. Fix a stochastic relation \( K : X \rightsquigarrow Y \), and a sub-probability measure \( \mu \in \mathcal{S}(X) \). Then \( \mu \otimes K \in \mathcal{S}(X \times Y) \) has a kind of natural converse: define \( \tau := \mathcal{S}(r)(\mu \otimes K) \), where \( r : X \times Y \to Y \times X \) switches components. Thus \( r[R] = R^\sim := \{ \langle y, x \rangle \mid \langle x, y \rangle \in R \} \), whenever \( R \subseteq X \times Y \) is a relation, so \( r \) produces the converse.

Because \( \tau \in \mathcal{S}(Y \times X) \), this measure is — according to Proposition 5.6.3 — representable through a stochastic relation \( K^\sim : Y \rightsquigarrow X \) and its projection \( \mathcal{S}(\pi_{Y \times X,Y})(\tau) \) upon writing

\[
\tau = \mathcal{S}(\pi_{Y \times X,Y})(\tau) \otimes K^\sim.
\]

Since \( \mathcal{S}(\pi_{Y \times X,Y})(\tau) = K^\sim(\mu) \) by Lemma 5.6.5, the definition of the converse of a stochastic relation now reads as follows.

**Definition 5.6.6** The \( \mu \)-converse \( K^\sim_\mu \) of the stochastic relation \( K \) with respect to the input probability \( \mu \) is defined by the equation

\[
\mathcal{S}(r)(\mu \otimes K) = K^\sim(\mu) \otimes K^\sim_\mu,
\]

where \( r : X \times Y \ni (x, y) \mapsto (y, x) \in X \times Y \) switches components.

It is remarked that by Proposition 4.2.6 the converse \( K^\sim_\mu \) always exists, and that it is unique \( \mu \)-almost everywhere. Since

\[
\mu(A) = (\mu \otimes K)(A \times Y) = (K^\sim(\mu) \otimes K^\sim_\mu)((Y \times A)^\sim)
\]

is true for the Borel set \( A \subseteq X \),

\[
\mu(A) = \int_X \int_Y K^\sim_\mu(A)(y) K(y)(dy) \mu(dx) = \int_X K^\sim_\mu(A) K^\sim(\mu)(dy),
\]

we infer that

\[
\mu = (K^\sim_\mu)^\bullet(\mu) = (K^\bullet K^\sim_\mu)^\bullet(\mu)
\]

holds. Hence the converse \( K^\sim_\mu \) solves the equation \( \mu = (K^\bullet T)^\bullet(\mu) \) for \( T \). This equation does, however, not determine the converse uniquely. This is so because it is an equation in terms of the Borel sets of \( X \), hence may only be carried over to the “strip” \( \{ A \times Y \mid A \in B(X) \} \) on the product \( X \times Y \). This is not enough to determine a measure on the entire product.

A **probabilistic interpretation** using regular conditional distributions may be given as follows: Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space, \( \zeta_i : \Omega \to X_i \) random variables with values in the Polish spaces \( X_i \) (\( i = 1, 2 \)). Let \( \mu \) be the marginal distribution of \( \zeta_1, \zeta_2 \), and let \( \mu_i \) be the marginal distribution of \( \zeta_i \). If \( \pi_i : X_1 \times X_2 \to X_i \) are the projections, then clearly \( \mu_i = \mathcal{S}(\pi_i)(\mu) \). \( K \) denotes the regular conditional distribution of \( \zeta_2 \) given \( \zeta_1 \), thus we have for the Borel sets \( A_i \subseteq X_i \)

\[
\mathbb{P}(\{ \omega \in \Omega \mid \zeta_1(\omega) \in A_1, \zeta_2(\omega) \in A_2 \}) = \mu(A_1 \times A_2) = \int_{A_1} K(x_1)(A_2) \mu_1(dx_1).
\]

158
We will show now that $K_{\mu_1}$ is the regular conditional distribution of $\zeta_1$ given $\zeta_2$. In fact, let $L$ be the latter distribution, then the definitions of $K$ and $L$, resp., imply

$$K^*(\mu_1) = \mu_2 \text{ and } L^*(\mu_2) = \mu_1.$$ 

Let $A_i \subseteq X_i$ be Borel sets, then

$$(K^*(\mu_1) \otimes L)(A_2 \times A_1) = \int_{A_2} L(x_2)(A_1) K^*(\mu_1)(dx_2)$$

$$= \int_{A_2} L(x_2)(A_1) \mu_2(dx_2)$$

$$= \int_{A_1} K(x_1)(A_2) \mu(dx_2)$$

$$= (\mu_1 \otimes K)(A_1 \times A_2).$$

Interpreting a stochastic relation as a regular conditional distribution of a random variable $\zeta_1$ given $\zeta_2$, its converse may be interpreted as the conditional distribution of $\zeta_2$ given $\zeta_1$. The start probability $\mu$ in the definition of $K_{\mu}$ is then interpreted as a marginal distribution. This is essentially the probabilistic setting for the definition of the converse in [2].

Returning to the general case, the defining equation for the converse is spelled out in terms of an integral:

$$\int_X K(x)(D^x) \mu(dx) = \int_Y K_{\mu}(y)(D_y) K^*(\mu)(dy).$$

This will be generalized and made use of later:

**Lemma 5.6.7** Let $f \in \mathcal{F}(X \times Y)$, then this identity holds:

$$\int_X \int_Y f(x, y) K(x)(dy) \mu(dx) = \int_Y \int_X f(x, y) K_{\mu}(y)(dx) K^*(\mu)(dy).$$

Thus the order of integration of $f$ may be interchanged, as in Fubini’s Theorem, but in contrast we need to adjust the measures used for integration (nevertheless it could be called *Fubinito’s Lemma*).

Some properties of forming the converse will be investigated now. We begin with an analogue of the property $R \bowtie \bowtie = R$ which holds for the set theoretic converse. Taking the initial distribution into account, this property is very similar for the probabilistic case.

**Proposition 5.6.8** If $K : X \sim Y$, and if $\mu \in \mathcal{G}(X)$, then $(K_{\mu})_{K^*(\mu)} = K$ holds everywhere except possibly on a set of $\mu$-measure zero.

**Proof** The stochastic relation $(K_{\mu})_{K^*(\mu)}$ is determined by the equation

$$(K^*(\mu) \otimes K_{\mu}) \bowtie = \eta \otimes (K_{\mu})_{K^*(\mu)}$$

with $\eta := K_{\mu}(K^*(\mu))$. The defining equation implies $\eta = \mu$, consequently $\mu \otimes K$ equals $\mu \otimes (K_{\mu})_{K^*(\mu)}$, as expected. \(\dashv\)
The question under what condition a stochastic relation may be represented as the converse of another relation is a little more difficult to answer than for the set-valued case. In view of the probabilistic interpretation using conditional distributions, however, the following solution arises naturally.

**Corollary 5.6.9** Let \( L : Y \leadsto X \) be a stochastic relation, and \( \mu \in \mathcal{S}(X) \). Then these conditions are equivalent:

1. \( \mu = L^*(\nu) \) for some \( \nu \in \mathcal{S}(Y) \),
2. \( L = K^*_\mu \) for some \( K : X \leadsto Y \).

Thus \( L : Y \leadsto X \) may be written in a variety of ways as the converse of a stochastic relations, viz., \( L = (K_\nu)_L^*(\nu) \) for an arbitrary \( \nu \in \mathcal{S}(Y) \) (where the relation \( X \leadsto Y \) depends on \( \nu \)). This is in marked contrast again to the set-theoretic case, where the converse of a relation is the relation itself, hence is uniquely determined. Compatibility of composition and forming the converse is an important property in the world of set-theoretic relations. In that case it is well known that \((R \circ S)\sim = S^\sim \circ R^\sim\) always holds (which might be called an anti-commutative law). The corresponding property for stochastic relations reads

**Proposition 5.6.10** Let \( K : X \leadsto Y \), \( L : Y \leadsto T \) be stochastic relations, and let \( \mu \in \mathcal{S}(X) \) be an initial distribution. Then \((K \circ L)\sim = L^\sim_K \circ (K^*_\mu) \sim\) holds.

**Proof** We will make use of Lemma 5.6.7 by showing that both relations have the same properties on measurable and bounded functions. Let \( f \in \mathcal{F}(X \times Z) \), then

\[
\int_{X \times Z} f \, d(\mu \otimes (K \circ L)) = \int_X \int_Z f(x, z) \, (K \circ L)(x)(dz) \, \mu(dx) \tag{5.1}
\]

\[
= \int_X \int_Y \int_Z f(x, z) \, L(y)(dz) \, K(x)(dy) \, \mu(dx) \tag{5.2}
\]

\[
= \int_Y \int_X \int_Z f(x, z) \, L(y)(dz) \, K^*_\mu(y)(dx) \, K^*(\mu)(dy) \tag{5.3}
\]

\[
= \int_Y \int_Z \int_X f(x, z) \, K^*_\mu(y)(dx) \, L(y)(dz) \, K^*(\mu)(dy) \tag{5.4}
\]

\[
= \int_Z \int_Y \int_X f(x, z) \, K^*_\mu(y)(dx) \, L^\sim_K \circ (K^*_\mu)(z)(dy) \tag{5.5}
\]

\[
= \int_Z \int_X \int_Y f(x, z) \, K^*_\mu(y)(dx) \, L^\sim_K \circ (K^*_\mu)(z)(dy) \, L^*(K^*(\mu))(dz) \tag{5.6}
\]

\[
= \int_Z \int_X \int_Y f(x, z) \, \left(L^\sim_K \circ (K^*_\mu)\right)(z)(dx) \, L^*(K^*(\mu))(dz). \tag{5.7}
\]

Equation (5.1) applies the definition of \( \mu \otimes (K \circ L) \) to the first integral. In equation (5.2) the definition of \( K \circ L \) is expanded, and in equation (5.3) Lemma 5.6.7 is applied to the two outermost integrals, similarly for equation (5.5). Fubini’s Theorem is used for interchanging integrals in equations (5.4) and (5.6). Finally, equation (5.7) applies the definition of the composition of kernels to \( L^\sim_K \circ (K^*_\mu) \) and \( K^*_\mu \).
On the other hand,
\[
\int_{X \times Z} f \, d(\mu \otimes (K \ast L)) = \int_{X} \int_{Z} f(x, z) \, (K \ast L)(z)(dx) \, \mu(dz)
\]
\[
= \int_{Z} \int_{X} f(x, z) \, (K \ast L)(z)(dx) \, L^\ast(K^\ast(\mu))(dz)
\]
is inferred from Lemma 5.6.7. Comparing the results established the claim. 

This is again a place to note algebraic similarities between set-theoretic and stochastic relations, but also to record exceptions. Take e.g. Schröder’s Cycle Rule

\[Q \ast R \subseteq S \iff Q^\ast \dot{\subseteq} R \iff \dot{S} \ast R^\ast \subseteq \dot{Q},\]

the bar denoting complementation ([90, 3.2 (xii)] or [12, Definition 3.1.1]). This rule is very helpful in practical applications [28], but it does not enjoy a direct counterpart for stochastic relations, since the respective notions of negation, and of containment do not carry over. —

If \(\mu(A) = 0\) for some Borel set \(A \subseteq X\), then \(K^\ast_{\mu^\ast}(y)(A) = 0\) holds \(K^\ast(\mu)-\text{almost everywhere on } Y\) (i.e., for all \(y \in Y\) outside a set of \(K^\ast(\mu)\)-measure zero). In fact, we can say more by having a closer look at the relationship between \(K^\ast_{\mu^\ast}\), \(K\) and \(\mu\). This leads to a rather surprising compactness result of the set of measures comprising the converse.

Recall that for \(\mu, \nu \in \mathcal{G}(X)\) the measure \(\nu\) is called absolutely continuous w. r. t. \(\mu\) iff for every measurable set \(A \subseteq X\) the implication \(\mu(A) = 0 \Rightarrow \nu(A) = 0\) holds; this is indicated by \(\nu \ll \mu\). It is well known [73] that \(\nu \ll \mu\) is equivalent to

\[\forall \varepsilon > 0 \exists \delta > 0 : [\mu(A) < \delta \Rightarrow \nu(A) < \varepsilon].\]

Absolute continuity is used for defining morphisms between probability spaces based on Polish spaces in [2, Definition 7.8] which in turn serves for defining the converse of a stochastic relation; we use it here for characterizing the measures comprising the converse. A subset \(M \subseteq \mathcal{G}(X)\) is accordingly called uniformly absolutely continuous w.r.t. \(\mu\) (indicated by \(M \ll \mu\)) iff given \(\varepsilon > 0\) there exists \(\delta > 0\) such that \(\sup_{\nu \in M} \nu(A) < \varepsilon\) whenever \(\mu(A) < \delta\) holds. It will be shown now that the set of measures constituting the converse is uniformly absolutely continuous except on a very small set:

**Proposition 5.6.11** Let \(K : X \leadsto Y\) be a stochastic relation, and \(\mu \in \mathcal{G}(X)\). Then for each version \(K^\ast_{\mu^\ast}\) of the converse of \(K\) with respect to \(\mu\) there exists a Borel set \(A \subseteq Y\) for which \(K(x)(A) = 0\) is true for \(\mu\)-almost all \(x \in X\), so that \(\{K^\ast_{\mu^\ast}(y) \mid y \notin A\} \ll \mu\) holds.

**Proof** 1. Let \(A \subseteq X\) be a Borel set with \(\mu(A) < \varepsilon\), then

\[\mu(A) \geq (\mu \otimes K)(A \times Y) = (K^\ast(\mu) \otimes K^\ast_{\mu^\ast})(Y \times A) = \int_{Y} K^\ast_{\mu^\ast}(y)(A) \, K^\ast(\mu)(dy),\]

thus there exists a measurable set \(N_{A} \subseteq Y\) such that \(K^\ast(\mu)(N_{A}) = 0\) and \(K^\ast_{\mu^\ast}(y)(A) < \varepsilon\) for all \(y \notin N_{A}\).

2. Since \(X\) is Polish, it is in particular second countable, thus there exists a countable base \(\mathcal{G}\) for the topology, and for the Borel sets. Put

\[N := \bigcup \{N_{G} \mid G \in \mathcal{G}\},\]
then $K^*(\mu)(N) = 0$.

3. Let $A$ be a Borel set with $\mu(A) < \varepsilon$, then there exists an open set $G \in \mathcal{T}$ with $A \subseteq G$ and $\mu(G) < \varepsilon$; this is so since finite measures on Polish spaces are regular [73, Theorem II.1.2]. Since $\mathcal{G}$ is a countable base for $\mathcal{T}$, we can cover $A$ by an increasing sequence $(G_n)_{n \in \mathbb{N}} \subseteq \mathcal{G}$ such that $\mu(\bigcup_{n \in \mathbb{N}} G_n) < \varepsilon$. Consequently, $K_\mu(y)(A) < \varepsilon$ for $y \notin N$. Thus $A := N$ is the desired set. \(\dashv\)

This implies that the set $\{K_\mu(y) \mid y \notin A\}$ is topologically not too large in the topology of weak convergence on $\mathcal{S}(X)$.

**Corollary 5.6.12** Let $X$ and $Y$ be Polish spaces, endow $\mathcal{S}(X)$ with the topology of weak convergence. Given $K : X \to Y$ and $\mu \in \mathcal{S}(X)$, then there exists a Borel set $A \subseteq Y$ with $K(x)(A) = 0$ for $\mu$-almost all $x \in X$, so that the set $\{K_\mu(y) \mid y \notin A\}$ is a relatively compact subset of $\mathcal{S}(X)$.

**Proof** Since finite measures on a Polish space are tight, we can find by Corollary A.3.4 for a given $\varepsilon > 0$ a compact set $C \subseteq X$ such that $\mu(X \setminus C) < \varepsilon$. The argumentation in the proof of Proposition 5.6.11 shows that

$$\sup_{y \notin A} K_\mu(x)(X \setminus C) < \varepsilon,$$

so that the set under consideration is uniformly tight. This implies the assertion by Prohorov’s Theorem A.3.3. \(\dashv\)

### 5.6.2 Preserving Bisimilarity

We will show that bisimilar relations give rise to bisimilar converses, so that bisimilarity is preserved under forming converses. We have to take into account, however, that forming the converse does not only depend on the relation itself, but that also an initial distribution is needed. Hence we extend the notion of bisimilarity to sub-probabilities as well by treating them as constant stochastic relations.

We have discussed different notions of bisimilarity with ties to congruences in this chapter. The variant that fits here best is 2-similarity, because domain and range of the mediating relation are part of the Cartesian product of the domain resp. range of the given relations, rather than being somewhat unrelated, abstractly given spaces. Thus bisimilar means in this section always 2-bisimilar.

**Definition 5.6.13** Let $X_1, X_2$ be Polish spaces with $\mu_i \in \mathcal{S}(X_i)$ ($i = 1, 2$). Then $\langle X_1, \mu_1 \rangle$ is said to be 2-bisimilar to $\langle X_2, \mu_2 \rangle$ iff there exists a subset $Z \subseteq X_1 \times X_2$ and $\zeta \in \mathcal{S}(Z)$ such that

1. $Z$ is a Borel subset of $X_1 \times X_2$,
2. $\mu_1 = \mathcal{S}(\pi_{Z,X_1})(\zeta)$ and $\mu_2 = \mathcal{S}(\pi_{Z,X_2})(\zeta)$.
3. there exists Borel sets $C_1 \subseteq X_1, C_2 \subseteq X_2$ with

$$\emptyset \neq Z \cap (C_1 \times X_2) = Z \cap (X_1 \times C_2) \neq Z.$$

$\langle Z, \zeta \rangle$ is said to mediate for $\langle X_1, \mu_1 \rangle$ and $\langle X_2, \mu_2 \rangle$.  

162
5.6 Case Study: The Converse of a Stochastic Relation

The first condition is quite necessary for otherwise it would be difficult to define a measure on \( Z \), the second one is just a translation of the requirement that the corresponding diagram should be commutative, and the third one postulates that there is a non-trivial common event, see section 5.4.

**Example 5.6.14** In the discrete setting, the mediating sup-probability measure may be represented as a matrix. In fact, let \( \{1, \ldots, n\}, \mu_1 \) and \( \{1, \ldots, m\}, \mu_2 \) be 2-bisimilar with mediating \( \langle Z, \zeta \rangle \). Then \( \zeta \) is represented as an \( n \times m \) matrix \( (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m} \) such that

1. \( 0 \leq a_{i,j} \leq 1 \),
2. for each \( i \), the sum \( \sum_{j=1}^{m} a_{i,j} \) equals \( \mu_1(i) \),
3. for each \( j \), the sum \( \sum_{i=1}^{n} a_{i,j} \) equals \( \mu_2(j) \).

The set \( Z \) is determined as the set of indices \( (i, j) \) for which \( a_{i,j} \neq 0 \).

Let \( X_1 = \{1, 2, 3\}, \mu_1 = [1/2, 1/4, 1/4] \) and \( X_2 = \{1, 2\}, \mu_2 = [3/8, 5/8] \). Then \( \langle Z, \zeta \rangle \) mediates between \( \langle X_1, \mu_1 \rangle \) and \( \langle X_2, \mu_2 \rangle \), where

\[
Z := \{(1, 2), (2, 1), (2, 2), (3, 1)\}
\]

and \( \zeta \) is given through the matrix

\[
\begin{bmatrix}
0 & 1/2 \\
1/8 & 1/4 \\
1/8 & 0
\end{bmatrix}
\]

\( \diamond \)

Bisimulations are maintained by forming products, and by transporting a measure through a stochastic relation, as we will see now:

**Proposition 5.6.15** Let \( K_i = \langle X_i, Y_i, K_i \rangle \) be 2-bisimilar Polish objects \( (i = 12) \) for which \( N : U \sim V \) mediates, and assume that \( \mu_i \in \mathcal{G}(X_i) \) such that \( \langle X_1, \mu_1 \rangle \) and \( \langle X_2, \mu_2 \rangle \) are 2-bisimilar with mediating \( \langle Z, \zeta \rangle \). Assume that \( Z \subseteq U \) holds, then

1. \( \langle Y_1, K_1^*(\mu_1) \rangle \) is 2-bisimilar to \( \langle Y_2, K_2^*(\mu_2) \rangle \) with mediating \( \langle V, N^*(\zeta) \rangle \),
2. \( \langle X_1 \times Y_1, \mu_1 \otimes K_1 \rangle \) is 2-bisimilar to \( \langle X_2 \times Y_2, \mu_2 \otimes K_2 \rangle \) with mediating \( \langle t[E], \mathcal{G}(t)(\zeta \otimes N) \rangle \), where \( E := Z \times V \) and \( t(x_1, x_2, y_1, y_2) := \langle x_1, y_1, x_2, y_2 \rangle \).

**Proof** 0. Because \( Z \subseteq U \), we know that for \( z \in Z \) the equality \( \pi_{Z,X_1}(z) = \pi_{U,X_1}(z) \) holds, so that \( K_1(\pi_{Z,X_1}(z)) = K_1(\pi_{U,X_1}(z)) = \mathcal{G}(\pi_{V,Y_1})(N(z)) \) is true; similarly for \( K_2 \).

1. For establishing 1, let \( f_1 \in \mathcal{F}(Y_1) \), then

\[
\int_{Y_1} f_1 \ dK_1^*(\mu_1) = \int_{X_1} \int_{Y_1} f_1(y_1) K_1(x_1)(dy_1) \mu_1(dx_1)
\]

\[
= \int_Z \int_{Y_1} f_1 \ dK_1(\pi_{Z,X_1}(z)) \ \zeta(dz)
\]

\[
= \int_Z \int_B (f_1 \circ \pi_{V,Y_1}) \ dN(z) \ \zeta(dz)
\]

\[
= \int_B f_1 \circ \pi_{V,Y_1} \ dN^*(\zeta).
\]

\[163\]
This implies $K_1^\bullet (\mu_1) = \mathcal{G} (\pi_{V,Y_1}) (N^\bullet (\zeta))$. In the same way, $K_2^\bullet (\mu_2) = \mathcal{G} (\pi_{V,Y_2}) (N^\bullet (\zeta))$ is established. This proves the first part of the assertion, because the $\sigma$-algebra of common events for $K_1$ and $K_2$ can be used for the common events of $(Y_1, K_1^\bullet (\mu_1))$ and $(Y_2, K_2^\bullet (\mu_2))$.

2. An argument very similar to the preceding one shows that for $f_1 \in F (X_1 \times Y_1)$ these equalities hold:

$$\int_{X_1 \times Y_1} f_1 \, d (\mu_1 \otimes K_1) = \int_{X_1} \int_{Y_1} f_1 (x_1, y_1) \, K_1 (x_1) (dy_1) \, \mu_1 (dx_1) = \int_{E} f_1 \, d (\mathcal{P} (\pi_{E,X_1 \times Y_1}) (\zeta \otimes N)).$$

A similar calculation shows for $f_2 \in F (X_2 \times Y_2)$ that

$$\int_{X_2 \times Y_2} f_2 \, d (\mu_2 \otimes K_2) = \int_{E} f_2 \, d (\mathcal{P} (\pi_{E,X_2 \times Y_2}) (\zeta \otimes N)).$$

This implies the assertion, since the isomorphism $t$ only serves to reorder variables.

The argumentation above shows that bisimilar relations and bisimilar initial distributions lead to bisimilar measures on the product. The process can be reversed: the idea is that disintegrating 2-bisimilar measures on a product leads to 2-bisimilar stochastic relations.

**Lemma 5.6.16** Let $X_i, Y_i$ be Polish spaces, $\mu_i \in \mathcal{G} (X_i \times Y_i)$ for $i = 1, 2$. Assume that $(X_1 \times Y_1, \mu_1)$ is 2-bisimilar to $(X_2 \times Y_2, \mu_2)$. Define the Polish objects $K_i := (X_i, Y_i, K_i)$ through the respective disintegrations of $\mu_i$ w.r.t $\mathcal{G} (\pi_{X_i \times Y_i}) (\mu_i)$. Then there exists a Polish object $M = (X_1 \times X_2, Y_1 \times Y_2, M)$ that mediates between $K_1$ and $K_2$.

**Proof.** Assume that $(E, \zeta)$ is mediating between $(X_1 \times Y_1, \mu_1)$ and $(X_2 \times Y_2, \mu_2)$. Put $E_0 := t (E), \zeta_0 := \mathcal{G} (t) (\zeta)$, where $t$ rearranges components, as in Proposition 5.6.15. Let $\gamma := \mathcal{G} (\pi_{E_0,X_1 \times X_2}) (\zeta_0) \in \mathcal{G} (X_1 \times X_2)$, and let $M'$ be the disintegration of $\zeta_0$ with respect to $\gamma$.

2. Let $\mathcal{G}_i$ be a countable generator for the $\sigma$-algebra on $X_i \times Y_i$, so that $\mathcal{G}_i$ is closed under finite intersections ($i = 1, 2$). Let $G \in \mathcal{G}_1$, then

$$\mu_1 (G) = \zeta_0 (G \times X_2 \times Y_2) = \int_{X_1 \times X_2} \mathcal{G} (\pi_{Y_1 \times Y_2,Y_1}) (M' (x_1, x_2)) (G_{x_1}) \, \gamma (d (x_1, x_2)).$$

and

$$\mu_1 (G) = \int_{X_1} K_1 (x_1) (G_{x_1}) \, \mathcal{G} (\pi_{X_1 \times Y_1,Y_1}) (\mu_1)$$

by the definition of $K_1$. Since

$$\mathcal{G} (\pi_{X_1 \times Y_1,Y_1}) (\mu_1) = \mathcal{G} (\pi_{X_1 \times Y_1,X_1}) (\mathcal{G} (\pi_{E_0,X_1 \times Y_1}) (\zeta_0)) = \mathcal{G} (\pi_{E_0,X_1}) (\zeta_0) = \mathcal{G} (\pi_{X_1 \times X_2,X_1}) (\gamma),$$

the latter integral may be expressed as

$$\mu_1 (G) = \int_{X_1 \times X_2} K_1 (x_1) (G_{x_1}) \, \gamma (d (x_1, x_2)).$$
Thus 

\[ A_G := \{ \langle x_1, x_2 \rangle \in X_1 \times X_2 \mid K_1(x_1)(G_{x_1}) \neq \mathcal{S}(\pi_{Y_1 \times Y_2}, Y_1)(M'(x_1, x_2))(G_{x_1}) \} \]

is a measurable subset of \( X_1 \times X_2 \) which has \( \gamma \)-measure 0. Put 

\[ A_1 := \bigcup \{ A_G \mid G \in G_1 \}, \]

then clearly \( \gamma(A_1) = 0 \), and \( K_1(x_1)(G_{x_1}) = \mathcal{S}(\pi_{Y_1 \times Y_2}, Y_1)(M'(x_1, x_2))(G_{x_1}) \) holds for all measurable subsets \( G \subseteq X_1 \times Y_1 \) whenever \( \langle x_1, x_2 \rangle \notin A_1 \). This is so since by the \( \pi \)-\( \lambda \)-Theorem A.1.1 a \( \cap \)-stable generator uniquely determines a finite measure, and since the equation above is true for all \( G \in G_1 \). In a similar way a measurable subset \( A_2 \) of \( X_1 \times X_2 \) can be found with \( \gamma(A_2) = 0 \), so that for \( \langle x_1, x_2 \rangle \notin A_2 \) and for all measurable subsets \( G \subseteq X_1 \times Y_2 \) the equality 

\[ K_2(x_2)(G_{x_2}) = \mathcal{S}(\pi_{Y_1 \times Y_2}, Y_1)(M'(x_1, x_2))(G_{x_2}) \]

holds.

3. Define \( M \) as \( M' \) outside \( A_1 \cup A_2 \), and set \( M(x_1, x_2) := K_1(x_1) \otimes K_2(x_2) \), for \( \langle x_1, x_2 \rangle \in A_1 \cup A_2 \), then \( M : X_1 \times X_2 \becomes Y_1 \times Y_2 \) has the desired properties. \( \dashv \)

Showing that bisimilarity is maintained when forming the converse is now an easy consequence:

**Proposition 5.6.17** Let \( K_i = \langle X_i, Y_i, K_i \rangle \) be 2-bisimilar Polish objects \( (i = 1, 2) \) between which \( N : U \becomes V \) mediates, and assume that \( \mu_i \in \mathcal{S}(X_i) \) such that \( \langle X_1, \mu_1 \rangle \) and \( \langle X_2, \mu_2 \rangle \) are 2-bisimilar with mediating \( \langle Z, \zeta \rangle \). Assume that \( Z \subseteq U \) holds. Then \( K_{1, \mu_1} \) is 2-bisimilar to \( K_{2, \mu_2} \).

**Proof** We know from Proposition 5.6.15 that \( \langle X_1 \times Y_1, \mu_1 \otimes K_1 \rangle \) and \( \langle X_2 \times Y_2, \mu_2 \otimes K_2 \rangle \) are 2-bisimilar. Bisimilarity is plainly not destroyed by interchanging coordinates. The assertion follows from Lemma 5.6.16, because the common events for \( \langle X_1, \mu_1 \rangle \) and \( \langle X_2, \mu_2 \rangle \) are also common events for the disintegrations. \( \dashv \)

### 5.7 Case Study: Simple Relations for Counting

The characterization of simple systems helps in analyzing the average behavior of algorithms by discussing two examples. The results are not new, the approach, however, is. Rutten [78, 79] shows how a stream calculus based on coinduction is used for counting, and hence for some aspects of the average case analysis of algorithms. This is made possible through the existence of final systems for the functor considered. By Corollary 5.5.7, the situation discussed here is different in that for the probabilistic case only a trivial final system exists, and it may be doubted whether this can be put to significant use.

Despite this somewhat restricted situation simple relations may be put to work; we will discuss the average case analysis of two algorithms and show how the continuous and the discrete case interact. This is done by constructing various relations that will be simple, and by deriving through Proposition 5.5.6 quantities that otherwise can be obtained only with difficulties. Denote in the discussion that follows by \( V_n \) the set of permutations on \( \{1, \ldots, n\} \).
5.7.1 Left to right maxima

Given an array \( a[1..n] \) of natural numbers, the following algorithm identifies the index \( m \) of the maximal element.

**Algorithm 5.7.1**

\[
m := 1; \\
\text{for } i := 2 \text{ to } n \text{ do} \\
\quad \text{if } a[m] < a[i] \text{ then } m := i; \text{ fi}; \\
\text{end for}; \]

The expected number of times the variable \( m \) changes its value in Algorithm 5.7.1 is asked for; Knuth discusses this algorithm and arrives at the result that this expectation equals \( H_n - 1 \), where \( H_n := \sum_{i=1}^{n} \frac{1}{i} \) is the \( n \)th harmonic number, provided the array has \( n \) mutually different components [53, Section 1.2.10]. To be more specific, he shows that the number \( p_{n,k} \) of permutations on \( \{1,\ldots,n\} \) for which the step in question is executed exactly \( k \) times equals

\[
p_{n,k} = \frac{1}{n!} \cdot \left[ \begin{array}{c} n \\ k \end{array} \right] \]

with \( \left[ \begin{array}{c} n \\ k \end{array} \right] \) as a Stirling number of the first kind [53, Section 1.2.10, Equation (9)]. These numbers are defined through

\[
z \cdot (z + 1) \cdot \ldots (z + n - 1) = \sum_{k} \left[ \begin{array}{c} n \\ k \end{array} \right] \cdot z^{k},
\]

they are interpreted combinatorially through cycles: \( \left[ \begin{array}{c} n \\ k \end{array} \right] \) is the number of ways to arrange \( n \) objects into \( k \) cycles, see [40, Section 6.1].

Define the stochastic relation \( K_n := (\{1,\ldots,n\}, 1, K_n) \) with \( K_n(k)(1) := p_{n,k} \), and assume that the values \( p_{n,0}, \ldots, p_{n,n} \) are mutually different (if they are not, factor). Then \( K_n \) is a simple relation, thus for each other relation \( K \) there exists at most one morphism into it. We want to compute the expected value when we have continuous data. Let \( (\Omega, A, \mathbb{P}) \) be a probability space, \( \zeta : \Omega \rightarrow [0,1]^n \) be a uniformly distributed random variable, and \( \tau : \Omega \rightarrow \mathbb{N} \) the number of times the value corresponding to \( m \) is changed. Thus if \( \omega \in \Omega \) is observed, the vector \( \zeta(\omega) \) is the input to the algorithm, \( \tau(\omega) = Z(\zeta(\omega)) \) counts the corresponding number, where \( Z \) is the function for counting. We are looking for the expected value \( E(\tau) \). Since \( \tau \) takes only discrete values, and since

\[
E(\tau) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(\tau = k),
\]

it is sufficient to compute the probability \( \mathbb{P}(\tau = k) \) that the random variable \( \tau \) has the value \( k \). Now put \( K(x)(1) := \mathbb{P}(\{\omega \in \Omega \mid \zeta(\omega) = Z(x)\}) \). Then the simplicity of \( K_n \) implies that defining \( K(x)(1) := p_{n,Z(x)} \) is the only way to define the stochastic relation \( K = ([0,1]^n, 1, K) \) making \( Z : K \rightarrow K_n \) a morphism. Consequently, \( \mathbb{P}(\tau = k) = p_{n,k} \), and \( E(\tau) = H_n - 1 \), as in the discrete case.

This illustrates how the transfer between discrete and continuous stochastic systems works: the behavior is known in the discrete case, and defining an appropriate simple system helps in transporting that knowledge to the continuous case. The example gives
5.7 Case Study: Simple Relations for Counting

insight into the relationship between discrete and continuous systems. The quantitative result is not new, however, and the central recurrence from which Knuth derives the expected value, viz.,

\[ p_{n,k} = \frac{1}{n} \cdot p_{n-1,k-1} + \frac{n-1}{n} \cdot p_{n-1,k}, \text{ and } p_{1,k} = \delta_{0,k} \]

can be easily derived directly for the continuous case.

5.7.2 Williams’ algorithm to construct heaps

Recall that a permutation \( p \in V_n \) is a heap iff \( p_{[i/2]} < p_i \) holds for each index \( i \) with \( 2 \leq i \leq n \). Heaps are usually represented through binary trees with node 1 as the root and node \( [i/2] \) as the father of node \( i \), so that the heap condition entails that each node has a label \( p_i \) which is larger than the label \( p_{[i/2]} \) of its father. Denote by \( H_n \) all elements of \( V_n \) that are heaps.

Let \( x \in [0,1]^n \) be a vector of \( n \) components taken from \( [0,1] \) which are mutually different, then \( \varphi_n(x) \in V_n \) is the permutation that arises from \( x \) by order statistics, i.e., if \( \varphi_n(x) = p \), then \( p_i = k \) iff \( x_i \) is the \( k \)th-largest component of \( x \). We will deal in the sequel with uniformly distributed elements of \( [0,1]^n \). Since equality of components happens only on a set of measure zero, those elements can be neglected, so that \( \varphi_n \) is defined almost everywhere on \( [0,1]^n \).

Assume that \( n \) has the binary representation \( /1b_{\nu-1} \ldots b_0/2 \), then the node

\[ t(n, \kappa) := /1b_{\nu-1} \ldots b_{\kappa-1}/2 \]

is called the special node on level \( \kappa \) [54, Exercise 5.2.3.20].

Now let

\[
\begin{align*}
S_{n+1,0} & := \{ p \in V_{n+1} \mid \varphi_n(p_1, \ldots, p_n) \in H_n, p_{n+1} > p_{t(n+1,1)} \}, \\
S_{n+1,\kappa} & := \{ p \in V_{n+1} \mid \varphi_n(p_1, \ldots, p_n) \in H_n, p_{t(n+1,\kappa)} > p_{n+1} > p_{t(n+1,\kappa+1)} \} \\
& \quad (1 \leq \kappa \leq \nu := \lfloor \log_2(n + 1) \rfloor), \\
S_{n+1,\nu} & := \{ p \in V_{n+1} \mid \varphi_n(p_1, \ldots, p_n) \in H_n, p_{t(n+1,\nu)} > p_{n+1} \}.
\end{align*}
\]

The task at hand is to count the number of elements in \( S_{n+1,0} \ldots S_{n+1,\nu} \). This is important for determining the average complexity of Williams’ algorithm to insert an element into a heap: Suppose \( p \in V_{n+1} \) is a permutation with \( n + 1 \) elements such that \( p_1, \ldots, p_n \) forms heap, then \( p_{n+1} \) is inserted into this heap according to Williams’ algorithm, which searches the path \( n + 1, [(n + 1)/2], \ldots, 1 \) that goes from node \( n + 1 \) to the root for the correct position of \( p_{n+1} \) and inserts it there, specifically:

**Algorithm 5.7.2**

\[
\begin{align*}
j & := n+1; \quad i := \lfloor j/2 \rfloor; \quad q := p_{n+1} \\
\text{while } (i > 0) \&\& (q < p_i) \text{ do} & \quad p_j := p_i; \quad j := i; \quad i := \lfloor i/2 \rfloor; \quad \od \\
\text{od}; \\
p_j & := q;
\end{align*}
\]
This algorithm may be used for iteratively building up a heap, it is one of the classics [97, 54]. Nevertheless, the average case analysis is surprisingly complicated [23]. We will show here through simple stochastic relations that probabilistic arguments help in counting permutations. Put

\[ W_n := \{ x \in [0, 1]^n \mid x \text{ is a heap} \}, \]
\[ G := W_n \times [0, 1]. \]

We will assume the inputs from \( W_n \) and from \([0, 1]\) are uniformly distributed with \( \lambda^n \) as Lebesgue measure on (the Borel sets of) \([0, 1]^n\). It is claimed for later use that

\[ \lambda^{n+1}(\{ x \in G \mid \varphi_{n+1}(x) = p \}) = \frac{\lambda^n(W_n)}{n+1} \]

independently of \( p \in V_{n+1} \). One first notes that the Change of Variable formula (Proposition 2.2.1) implies that

\[ \lambda^{n+1}(\{ x \in G \mid \varphi_{n+1}(x) = p \}) = \lambda^{n+1}(\{ x \in G \mid \varphi_{n+1}(x) = p' \}) \]

with \( p' \) as the element of \( V_{n+1} \) the first \( n \) components of which form a heap, and \( p'_{n+1} < p'_1 \). This is so since the Jacobian of a permutation for the coordinates equals 1. Thus we obtain

\[ \lambda^{n+1}(\{ x \in G \mid \varphi_{n+1}(x) = p \}) = \lambda^{n+1}(\{ x \in G \mid \varphi_{n+1}(x) = p' \}) = \frac{\lambda^n(W_n)}{n+1} \]

Equation (†) is Fubini’s Theorem on product integration, equation (‡) is again the Change of Variable formula, since the transformation

\[ (y_1, \ldots, y_n) \mapsto ((1 - x) \cdot y_1 + x, \ldots, (1 - x) \cdot y_n + x) \]

which maps \( \{ y \in [0, 1]^n \mid y \text{ is a heap} \} \) bijectively to \( \{ y \in [0, 1]^n \mid y \text{ is a heap}, x < y_1 \} \) has the Jacobian \( (1 - x)^n \).

Now let \( g(n, i) \) be the number of nodes in the subtree rooted at node \( i \), then Knuth [54, Equation 5.2.3-14] shows that \( g(n, /b_{\nu-1} \ldots b_{\kappa}/2) = /1b_{\kappa-1} \ldots b_{\nu}/2 \), and that, if node \( i \) is on level \( \kappa \),

\[ g(n, i) = \begin{cases} 2^{\nu-\kappa} - 1 & \text{if } i \text{ is a right node,} \\ 2^{\nu-\kappa+1} - 1 & \text{if } i \text{ is a left node} \end{cases} \]

where right and left are relative to the node’s position with respect to the special node on that level.

It follows from [54, Equation 5.2.3-16, Exercise 5.1.4-20] that for the number \( h_n \) of heaps on \( n \) elements

\[ h_n = \prod_{i=1}^{n} g(n, i) \]
Proposition 5.7.3 If \( n + 1 \) has the binary representation \( /1c_{v-1} \ldots c_0/2 \), then

\[
\chi_n(\mathcal{S}(\kappa)) = \frac{\chi_n}{/1c_{\kappa-1} \ldots c_0/2} \cdot \prod_{j=\kappa+1}^{\nu} \frac{/1c_{j-1} \ldots c_0/2 - 1}{/1c_{j-1} \ldots c_0/2}.
\]

Now define the stochastic relation \( K = (G, \emptyset, K) \) through \( K(x)(\emptyset) := \chi_n(\mathcal{S}(Z(x))) \), and put \( K' := (\{0, \ldots, \lfloor \log_2(n+1) \rfloor\}, \emptyset, K') \) with

\[
K'(\kappa)(\emptyset) := \frac{\chi_n}{n+1} \cdot |\mathcal{S}_{n+1, \kappa}|.
\]

We claim that \( K(x)(\emptyset) = K'(Z(x))(\emptyset) \) holds for each \( x \in G \), rendering \( K \to K' \). In fact,

\[
|\mathcal{S}_{n+1, \kappa}| = \sum_{p \in \mathcal{S}_{n+1, \kappa}} 1 = \frac{n+1}{\chi_n} \cdot \sum_{p \in \mathcal{S}_{n+1, \kappa}} \chi_n^{n+1}(\{x \in \mathcal{S}(\kappa) \mid \varphi_{n+1}(x) = p\})
\]

\[
= \frac{n+1}{\chi_n} \cdot h_n \cdot \chi_n^{n+1}(\{x \in G \mid \varphi_{n+1}(x) \in \mathcal{S}_{n+1, \kappa}\})
\]

\[
= \frac{n+1}{\chi_n} \cdot h_n \cdot \chi_n^{n+1}(\mathcal{S}(\kappa)).
\]

This is a cornerstone for establishing

Proposition 5.7.4 The number \( |\mathcal{S}_{n+1, \kappa}| \) of permutations \( p \) on \( \{1, \ldots, n+1\} \) such that \( p_1, \ldots, p_{n+1} \) is a heap, and \( p_{n+1} \) climbs up \( \kappa \) levels during Williams’ algorithm equals

\[
\frac{(n+1) \cdot h_n}{/1c_{\kappa-1} \ldots c_0/2} \cdot \prod_{j=\kappa+1}^{\nu} \frac{/1c_{j-1} \ldots c_0/2 - 1}{/1c_{j-1} \ldots c_0/2},
\]

assuming that \( n + 1 \) has the binary representation \( /1c_{v-1} \ldots c_0/2 \).
Proof We assume again without loss of generality that the numbers in question are mutually different. Then they represent the only way of defining the simple system $K'$ in such a way that $Z$ becomes a morphism. $\dashv$

This explicit result is apparently new. It should be noted, however, that results of this kind cannot be used for exploiting the average complexity of Williams' algorithm. It makes essential use of the fact that the underlying probability is uniform, but it is well known that uniform distribution is destroyed by inserting an element into a heap through this algorithm (or removing an element from it using Floyd's).

5.8 Bibliographic Notes

Bisimilarity. Bisimilarity is introduced essentially as a span of morphisms [50, 77, 25]. For coalgebras based on the category of sets, this definition agrees with the one through relations, originally given by Milner, see [77]. In [21] the authors call a bisimulation what we introduced as a congruence, albeit that paper restricts itself to labelled Markov transition systems, thus technically to families of stochastic relations $S \rightsquigarrow S$ for some state space $S$. It seems conceptually to be clearer to distinguish spans of morphisms from equivalence relations, thus we make this distinction here.

The close relationships between bisimulations and certain equivalence relations have been known since at least the Hennessy-Milner Theorem [44]. Having a closer look at the relation that is defined there, it is evident that the relation defined through

$$s \equiv s' \iff \forall \phi : [s \models \phi \iff s' \models \phi]$$

is countably generated. Once one can show in a probabilistic interpretation of modal logics that the sets of states for which a formula is valid is measurable, the relation is recognized as smooth, so the tools from the theory of Borel sets [51, 88, 5] developed for countably generated equivalence relations become available; this is exactly how we will proceed in section 6.1. The observation that factoring an analytic space through a smooth equivalence relation will yield an analytic space again is the reason that analytic spaces figure prominently in this development, on the other hand it shows upon factoring that these relations will not lead to bizarre spaces that are technically difficult to grasp.

The Converse. Abramsky, Blute, and Panangaden [2] investigate the category $\text{PRel}$ of probability spaces, hereby introducing the converse of a probabilistic relation as we do through the product measure [2, Section 7]. The process by which they arrive at this construction is quite similar to disintegration, as proposed here but makes heavier use of absolute continuity (in fact, morphisms in $\text{PRel}$ use absolute continuity in a crucial way). The argumentation that has been used in the present discussion seems to be closer to the set-theoretic case by looking at what happens when we compute the probability for a converse relation. Further investigations of the converse do not include the anti-commutative law Proposition 5.6.10. This is probably due to the fact that integration technique are directly used in the present paper (while [2] prefers arguing using absolute continuity, and consequently, with the Radon-Nikodym Theorem).

The analogy between set-theoretic and stochastic relations is like a central thread to many investigations in this area, since it is sometimes annoying, sometimes exciting to see that constructions that can be carried out without great difficulties for relations can
be done only with great effort for the probabilistic case. An example is given through the converse, another one is apparent when studying pullbacks. The monograph [37] is a general approach to fit relations under one roof.

**Simple Systems.** Simple systems are of course a topic in algebra [58]. The work in coalgebras for which [77] stands as a representative has given a tight relationship between simplicity and certain forms of bisimulations. This is not only for reasons to better explore the structure of coalgebras but also because final systems are at the basis of the proof principle of coinduction. It is based on the observation that a final system has exactly one morphism going into it, so it constructs its argument usually by constructing scenarios from which the desired properties are inferred through uniqueness. This requires the existence of a final system, which in turn is sometimes a rather complicated business, and usually the underlying functor must at least preserve weak pullbacks. This situation is analyzed in a paper by Gumm and Schröder for the category of sets, and it is shown that the functor governing the coalgebra preserves kernel pairs iff every congruence is a bisimulation [41, Theorem 5.7]. It is interesting to compare this statement with Proposition 4.3.6 and Proposition 5.4.2 for the sub-probability functor.

In [70] the probability functor $P$ is considered on measurable spaces that are endowed with a initial $\sigma$-algebra related to the weak-*-$\sigma$-algebra. Given a discrete space $I$, final coalgebras for a functor derived from $P$ yielding a type space over category $\text{Meas}^I$ are discussed. Besides establishing the existence of a final coalgebra for the functor through satisfied theories, the main result states that functors polynomial in the original type functor have final coalgebras as well. This result is extended in [93] using the final sequence of the functor under consideration, where the method is shown to work for other, set-valued functors as well. In [17] coalgebraic simulation is considered, and one of the application areas for the discussion are probabilistic transition systems. They are modelled as coalgebras for the functor that assigns each set its discrete probability measures. All this shows that special assumptions and constructions are needed for securing the existence of a final system that is related with the probability functor even over finite spaces.
Chapter 6

Interpreting Modal and Temporal Logics

Contents

6.1 Modal Logics ................................................. 175
  6.1.1 Examples ............................................. 177
  6.1.2 Refinements .......................................... 180
  6.1.3 Bisimulations for Kripke Models ..................... 183

6.2 Infinite Paths for Interpreting Temporal Logics .......... 189
  6.2.1 Probabilities for Paths .............................. 189
  6.2.2 pCTL* .................................................. 190
  6.2.3 CSL ..................................................... 191

6.3 Bisimulations for CSL .................................... 201
  6.3.1 Definition and Properties of ρF ...................... 202
  6.3.2 Closure Operations .................................. 207
  6.3.3 F-Bisimulations .................................... 208

6.4 Bibliographic Notes ....................................... 212

Consider the simple modal logic the formulas of which are given through

\[ \varphi ::= \top \mid p \mid \varphi_1 \land \varphi_2 \mid \Diamond \varphi, \]

where \( p \in P \) with \( P \) a set of atomic propositions. A Kripke model \( R = (S, R, V) \) is given by a set \( S \) of states (the possible worlds), a relation \( R \subseteq S \times S \) and a map \( V : P \to \mathcal{P} \omega(S) \). \( V(p) \) indicates for an atomic proposition in which worlds it holds. Given \( s \in S \), we say that formula \( \Diamond \varphi \) holds in \( s \) iff we can find \( s' \in S \) with \( \langle s, s' \rangle \in R \) such that \( \varphi \) holds in \( s' \). Thus \[ [\Diamond \varphi] = (\exists R)([[\varphi]]) , \] where as usual \([\varphi]\) is the set of all states in which formula \( \varphi \) holds.

A probabilistic counterpart for Kripke models cannot restrict itself to stating that a formula holds (or that it doesn’t) but should provide quantitative arguments: we want to know the probability with which a formula is true. Thus we want to know the probability for \( \Diamond \varphi \) to hold, subject to \( \varphi \) being true. Hence we replace the relation \( R \subseteq S \times S \) by a stochastic relation \( K : S \rightsquigarrow S \), and the single diamond \( \Diamond \) by a whole family \( (\Diamond_q)_{0 \leq q \leq 1} \).
indicating the degree to which $\Diamond \varphi$ holds, and state formally that in world $s \in S$ the formula $\Diamond_q \varphi$ is true iff $K(s)([\varphi]) \geq q$ holds. This is the faithful probabilistic counterpart to the model above.

This chapter studies probabilistic interpretations of modal and temporal logics. It does not cost much more to replace the very simple modal logic from above by a more comfortable one which commands a collection of modal operators of arbitrary arity, and which includes negation. We will define in section 6.1 nondeterministic and stochastic Kripke models, give examples for some well-known logics, and we will compare the nondeterministic and the stochastic approach. A stochastic Kripke model is seen as a quantitative model, while a nondeterministic one obviously stresses the qualitative character. Consequently it is sensible to relate nondeterministic to stochastic models through a refinement relation: if a formula holds in the stochastic model with probability one, then it should hold as well in the nondeterministic one. This relation between models is investigated, and we show under which conditions a nondeterministic model can be refined stochastically. Not surprisingly, selection theorems for set-valued maps enter the argumentation.

But the really interesting topic is bisimilarity — under what conditions are two stochastic Kripke models bisimilar? We show that the Hennessy-Milner Theorem provides an answer in this scenario as well: bisimilarity and identical theories are equivalent. This requires, however, a careful discussion of morphisms and the bisimulations associated with them. At this point we harvest from the investment we undertook in discussing bisimulations from a very general and abstract point of view, since we can just stretch out our hands and pluck the results, once the scenario is set up properly. This, then, gives a rather general result for stochastic Kripke models, and the Bibliographic Notes at the end of the present chapter will put things into the proper context.

We will turn our attention then to temporal logics and propose interpretations for the well known logics $\text{pCTL}^*$ and $\text{CSL}$. The important point to be made is to show first how to interpret path formulas, i.e. formulas the validity of which depends on an infinite path. This requires some measure-theoretic preparations, since we have to build up a probability that works on infinite paths from the probabilities for just one step being performed. The main tool to use will be the projective limit of a suitable projective system.

The difference between $\text{pCTL}^*$ and $\text{CSL}$ is among others the explicit incorporation of residence times into the latter, and we will show that the tools we collected from smooth equivalence relations and from congruences can be put to good use for investigating subsets of all formulas that govern the behavior of a model on all formulas. Putting it less cryptic, we take a subset $F \subseteq \mathcal{L}_P$ of all formulas, and ask the question, under what conditions the equivalence relation

$$s \equiv_F s' \iff \forall \phi \in F : [s \models \phi \iff s' \models \phi]$$

equals the equivalence relation $\equiv_{L_P}$ (with $P$ again as the set of atomic formulas). It is clear that the answer is interesting for model checking, since such a set $F$ of formulas eases the task of a model checker tremendously. We will look into the case that $F = P$, identifying under which conditions the simplest possible case will do. Surprisingly, it turns out that the invariant Borel sets for the smooth relation $\equiv_F$ will play a leading rôle.
The stochastic interpretation of CSL proposed here is new. In contrast to the traditional approach that starts from a rate function, this approach assumes only the stochastic independence of state changes and residence times. This permits understanding the interpretation as a stochastic relation, enabling the use of the tools developed for investigating these relations.

The main contributions of this approach to CSL and the understanding of stochastic logics in general lie in investigation of bisimulations for these logics as congruences together with the development of criteria for the equivalence of different notions of bisimulations. Another point worth emphasizing is that this approach permits the formulation of a general approach for the investigation of bisimulations for this type of logic through the theory of congruences for stochastic relations (which, in turn, is developed further). These logics all have their roots in approaches to model checking, thus they have a pronounced practical side. For the logic at hand this means that computational issues are important, and it becomes evident that structural properties need to be looked at not only for their own interest. The results for CSL entail such practical considerations.

6.1 Modal Logics

We have established a criterion for bisimilarity through equivalent congruences and discussed bisimilarity in terms of isomorphic factor spaces, see in particular section 5.3. We will now apply this to modal logic. This section defines the logic we will be working with, and Kripke models are defined in their usual nondeterministic and their stochastic versions, together with their satisfaction relation. In section 6.1.1 some examples are given in order to exhibit probabilistic models for specific logics, and we relate in section 6.1.2 nondeterministic to stochastic interpretations by introducing probabilistic refinements.

Let $P$ be a countable set of propositional letters which is fixed throughout, $O \neq \emptyset$ is a set of modal operators. Following [11], $\tau = (O, \rho)$ is called a modal similarity type iff $O \neq \emptyset$, and if $\rho : O \to \mathbb{N}$ is a map, assigning each modal operator $\triangle$ its arity $\rho(\triangle) \geq 1$. We will not deal with modal operators of arity zero, since they do not have to be dealt with as modal constants in an interpretation. The similarity type $\tau$ will be fixed.

We define three modal languages based on $\tau$ and $P$. The formulas of the basic modal language $\text{Mod}_b(\tau, P)$ are given by the syntax

$$\varphi ::= p \mid \top \mid \varphi_1 \land \varphi_2 \mid \neg \varphi \mid \triangle(\varphi_1, \ldots, \varphi_{\rho(\triangle)})$$

where $p \in P$. If we have $O' = \{\Diamond\}$ with $\rho(\Diamond) = 1$, we obtain the formulas of the basic modal language with negation. Omitting negation in $\text{Mod}_b(\tau, P)$ defines the formulas in the negation free basic modal language $\text{Mod}_1(\tau, P)$. Finally the extended modal language $\text{Mod}_e(\tau, P)$ is defined through the syntax

$$\varphi ::= p \mid \top \mid \varphi_1 \land \varphi_2 \mid \neg \varphi \mid \triangle_q(\varphi_1, \ldots, \varphi_{\rho(\triangle)})$$

where $q \in \mathbb{Q} \cap [0, 1]$ is a rational number, and $p \in P$ is a propositional letter. Again, if we deal with $O = O'$, then we get an entire line of new formulas through $(\Diamond_q)_{q \in \mathbb{Q} \cap [0, 1]}$.

A nondeterministic $\tau$-Kripke model $\mathcal{R} = (S, R_{\tau}, V)$ consists of a state space $S$, a family $R_{\tau} = ((R_\triangle)_{\triangle \in O})$ of set valued maps $R_\triangle : S \to \mathcal{P}(S^{\rho(\triangle)})$ and a set valued map $V : P \to \mathcal{P}(S)$.
The satisfaction relation $\models$ for a nondeterministic $\tau$-Kripke model $\mathcal{R}$ is defined as usual for $\mathfrak{Mod}_b(\tau, P)$:

- $\mathcal{R}, s \models p \iff s \in V(p)$
- $\mathcal{R}, s \models \neg \varphi \iff \mathcal{R}, s \not\models \varphi$
- $\mathcal{R}, s \models \varphi_1 \land \varphi_2 \iff \mathcal{R}, s \models \varphi_1$ and $\mathcal{R}, s \models \varphi_2$
- $\mathcal{R}, s \models \triangle(\varphi_1, \ldots, \varphi_{\rho(\triangle)}) \iff \exists (s_1, \ldots, s_{\rho(\triangle)}) \in R_\triangle(s) : \mathcal{R}, s_i \models \varphi_i$ for $1 \leq i \leq \rho(\triangle)$.

Denote by

$$[\varphi]_\mathcal{R} := \{ s \in S | \mathcal{R}, s \models \varphi \}$$

the set of states for which formula $\varphi$ is valid, and by

$$Th_\mathcal{R}(s) := \{ \varphi \in \mathfrak{Mod}_b(\tau, P) | \mathcal{R}, s \models \varphi \}$$

the theory of state $s$ in $\mathcal{R}$.

An easy calculation shows that

$$\mathcal{R}, s \models \triangle(\varphi_1, \ldots, \varphi_{\rho(\triangle)}) \iff R_\triangle(s) \cap [\varphi_1]_\mathcal{R} \times \ldots \times [\varphi_{\rho(\triangle)}]_\mathcal{R} \neq \emptyset \iff s \in (\exists R_\triangle)([\varphi_1]_\mathcal{R} \times \ldots \times [\varphi_{\rho(\triangle)}]_\mathcal{R}).$$

In analogy, a stochastic $\tau$-Kripke model $\mathcal{K} = (S, K_\tau, V)$ has a state space $S$ which is endowed with a $\sigma$-algebra $\mathcal{A}$, a family $K_\tau = (K_\triangle)_{\triangle \in \Theta}$ of stochastic relations $K_\triangle : S \rightsquigarrow S^{\rho(\triangle)}$ and a set valued map $V : P \rightarrow \mathcal{A}$. We will always assume that $S$ is a Polish space, and that the $\sigma$-algebra are the Borel sets.

The interpretation of formulas in $\mathfrak{Mod}_s(\tau, P)$ for a stochastic $\tau$-Kripke model $\mathcal{K}$ is fairly straightforward, the interesting case arising when a modal operator is involved:

$$\mathcal{K}, s \models \triangle_q(\varphi_1, \ldots, \varphi_{\rho(\triangle)})$$

holds iff there exists measurable subsets $A_1, \ldots, A_{\rho(\triangle)}$ of $S$ such that $\mathcal{K}, s_i \models \varphi_i$ holds for all $s_i \in A_i$ for $1 \leq i \leq \rho(\triangle)$, and

$$K_\triangle(s)(A_1 \times \ldots \times A_{\rho(\triangle)}) \geq q.$$

Arguing from the point of view of state transition systems, this interpretation of validity reflects that upon the move indicated by $\triangle_q$, a state $s$ satisfies $\triangle_q(\varphi_1, \ldots, \varphi_{\rho(\triangle)})$ iff we can find states $s_i$ satisfying $\varphi_i$ with a $K_\triangle$-probability exceeding $q$. Note that the usual operators $\triangle$ and $\triangledown$ are replaced by a whole spectrum of operators $\triangle_q$ which permit a finer and probabilistically more adequate notion of satisfaction.

Again, let $[\varphi]_\mathcal{K}$ be the set of all states for which $\varphi \in \mathfrak{Mod}_s(\tau, P)$ is satisfied under $\mathcal{K}$, and

$$Th_\mathcal{K}(s) := \{ \varphi \in \mathfrak{Mod}_s(\tau, P) | \mathcal{K}, s \models \varphi \}$$

the theory for state $s \in S$.

It turns out that the sets $[\varphi]_\mathcal{K}$ are measurable, so that they may be used as arguments for the stochastic relations we are working with:
**Lemma 6.1.1** \( [\varphi]_K \) is a measurable subset of \( S \) for each \( \varphi \in \text{Mod}_S(\tau, P) \).

**Proof** The proof proceeds by induction on \( \varphi \). If \( \varphi = p \in P \), then \( [\varphi]_K = V(p) \) holds, which is measurable by assumption. Since the measurable sets are closed under complementation and intersection, the only interesting case is again the one in which a modal operator is involved. Because
\[
[\triangle_q(\varphi_1, \ldots, \varphi_{\rho(\triangle)})]_K = \{ s \in S | K_\triangle(s)([\varphi_1]_K \times \ldots \times [\varphi_{\rho(\triangle)}]_K) \geq q \},
\]
the assertion follows from the induction hypothesis and the fact that \( K_\triangle \) is a stochastic relation. \( \dashv \)

As in the case of stochastic relations we need to exclude trivial cases:

**Definition 6.1.2** A \( \tau \)-Kripke model \( K \) with state space \( S \) is called degenerate iff \( [\varphi]_K = S \) or \( [\varphi]_K = \emptyset \) holds for each formula \( \varphi \in \text{Mod}_S(\tau, P) \).

Hence a degenerate model does usually not carry useful information. The restriction is quite similar to not permitting the universal relation as a part of a congruence, and of requesting the existence of non-trivial common events for bisimulations. We will see that these constraints are closely related.

### 6.1.1 Examples

We show how some popular logics may be interpreted through Kripke models, indicating that specific logics require specific probabilistic arguments. But first we indicate that each stochastic relation may be “trained” to interpret a modal logic simply by interpreting the subformulas of a compound formula as stochastically independent. Then we introduce the well-known logic associated with labelled transition systems. This example is of historic significance, given the seminal work of Larsen and Skou [59]. It is shown also that the basic temporal language may be interpreted stochastically by reversing a relation. Arrow logic as a popular logic modelling simple programming constructs is interpreted through a simple transformation of a distribution. The last example leaves the realm of modal logics somewhat by tackling a logic that is used for model checking. It will be shown how a stochastic relation generates path probabilities (through an inverse limit construction), and that this can be made use of for a probabilistic interpretation of a simple kind of tree logic. In presenting these examples we follow essentially the representation of the respective logics in [11].

A stochastic relation on the state space induces a stochastic \( \tau \)-Kripke model. This is illustrated through the following example.

**Example 6.1.3** Let \( K : S \rightsquigarrow S \) be a stochastic relation on the state space \( S \), and define for \( s \in S \) and for the modal operator \( \triangle \)
\[
K_\triangle(s) := \bigotimes_{i=1}^{\rho(\triangle)} K(s),
\]
then \( K_\triangle : S \rightsquigarrow S^{\rho(\triangle)} \) is a stochastic relation. Let \( V : P \rightarrow B(S) \), then
\[
K_{K,V} := (S, (K_\triangle)_{\triangle \in O}, V)
\]
is a stochastic $\tau$-Kripke model such that

$$K_{K,V,s} \models \Delta_q(\varphi_1, \ldots, \varphi_{\rho(\Delta)}) \iff K(s)([\varphi_1]_{K,V}) \cdots K(s)([\varphi_{\rho(\Delta)}]_{K,V}) \geq q.$$ 

Thus the arguments to each modal operator are stochastically independent. ◇

**Example 6.1.4** Suppose that $L$ is a countable alphabet of actions. Each action $a \in L$ is associated with a binary modal operator $\langle a \rangle$, so put $\tau := (O, \rho)$ with $O := \{ \langle a \rangle | a \in L \}$ and $\rho(\langle a \rangle) := 1$.

A nondeterministic $\tau$-Kripke model is based on a labelled transition system $(S, (\rightarrow_a)_{a \in L})$ which associates a binary relation $\rightarrow_a \subseteq S \times S$ with each action $a$. Thus

$$s \models \langle a \rangle \varphi \iff \exists s' : s \rightarrow_a s' \wedge s' \models \varphi.$$ 

A stochastic $\tau$-model is based on a labelled Markov transition system $(S, (k_a)_{a \in L})$ which associates with each action $a$ a stochastic relation $k_a : S \rightsquigarrow S$. Thus

$$s \models \langle a \rangle q \varphi \iff k_a(s)([\varphi]) \geq q,$$

hence making a transition is replaced by a probability with which a transition can happen.

Variants of the logic $\mathfrak{Mod}_\tau(\tau, P)$ with $P = \emptyset$ were investigated in the literature by Larsen and Skou, and by Desharnais, Edalat and Panangaden with a reference to the logic investigated by Hennessy and Milner [44]; we refer to them also as *Hennessy-Milner logic*. ◇

**Example 6.1.5** The basic temporal language has two unary modal operators $F$ (forward) and $B$ (backward), so that $O = \{ F, B \}$. A nondeterministic $\tau$-Kripke model interprets the forward operator $F$ through a relation $R \subseteq S \times S$ and the backward operator $B$ through the converse $R^\leftarrow$ of relation $R$, thus $R^\leftarrow := \{ \langle s', s \rangle | \langle s, s' \rangle \in R \}$. Consequently, we have

$$s \models B \varphi \iff \exists t \in S : \langle t, s \rangle \in R \land t \models \varphi.$$ 

A probabilistic interpretation interprets $F$ through a stochastic relation $K : S \rightsquigarrow S$, so that

$$s \models F_q \varphi \iff K(s)([\varphi]) \geq q.$$ 

The backward operator $B$ is interpreted through the converse $K^\leftarrow : S \rightsquigarrow S$, provided the state space $S$ is Polish and an initial probability $\mu$ is given. It was shown in section 5.6 that the converse $K^\leftarrow$ of stochastic relation $K$ given $\mu$ is the stochastic relation $L : S \rightsquigarrow S$ such that

$$\int_S K(s)(B_s) \mu(ds) = \int_S L(s')(B^s') \mu(ds')$$

holds for each Borel set $B \subseteq S \times S$. We know from Lemma 5.6.5 that for Polish $S$ the converse relation exists. Thus

$$s \models B_q \varphi \iff K^\leftarrow(s)([\varphi]) \geq q.$$ 

An easy calculation shows that

$$s \models B_1 F_1 \varphi \iff K^\leftarrow(s) \{ s' | K(s')([\varphi]) = 1 \} = 1 \iff \int_S K(s')([\varphi]) K^\leftarrow(s)(ds') = 1.$$
Note that the definition of the converse requires an initial probability (this is intuitively clear: if the probability for a backward running process is described, one has to say where to start). It is also noteworthy that a topological assumption has been made; if the state space is not a Polish space, then the technical arguments permitting the definition of the converse are not available.

**Example 6.1.6** Arrow logic has three modal operators modelling reversal, composition and \( \text{\&op}\), resp., thus \( O = \{1, \otimes, \circ\} \). with respective arities \( \rho(1) = 0, \rho(\otimes) = 1, \rho(\circ) = 2 \). The usual interpretation of arrow logic is done over a world of pairs, so the base state space is \( S \times S \) for some \( S \), with associated relations

\[
R_1 = \{\langle s, s \rangle \mid s \in S\} \\
R_\otimes = \{\langle \langle s_0, s_1 \rangle, \langle s_1, s_0 \rangle \rangle \mid s_0, s_1 \in S\} \\
R_\circ = \{\langle \langle s_0, s_1 \rangle, \langle s_0, s \rangle, \langle s, s_1 \rangle \rangle \mid s, s_0, s_1 \in S\}.
\]

Thus e.g.

\[
\langle s, s' \rangle \models \phi \circ \psi \iff \exists s_0 : \langle s, s_0 \rangle \models \phi \land \langle s_0, s' \rangle \models \psi
\]

and

\[
\langle s, s' \rangle \models \otimes \psi \iff \langle s', s \rangle \models \phi.
\]

Now assume again that \( S \) is a Polish space, and let \( \mu \in \mathcal{G}(S) \) be a sub-probability. Put for \( A \in \mathcal{B}(S \times S) \)

\[
\hat{\mu}(A) := \mu(\{s \in S \mid \langle s, s \rangle \in A\}),
\]

thus \( \hat{\mu} \) transports a Borel set in \( S \) to a Borel set in the diagonal of \( S \times S \). Interpret the composition operator \( \circ_q \) through the stochastic relation

\[
K_\circ(s, s') := \delta_s \otimes \hat{\mu} \otimes \delta_{s'}.
\]

Note that the operator \( \otimes \) is somewhat overloaded: it denotes the modal operator for reversal, and the product operator for measures. The context should make it clear which version is meant.

We obtain then

\[
K_\circ(s, s')(\llbracket \phi \rrbracket \times \llbracket \psi \rrbracket) = (\delta_s \otimes \hat{\mu} \otimes \delta_{s'})(\llbracket \phi \rrbracket \times \llbracket \psi \rrbracket)
\]

\[
= \hat{\mu}(\{s_1, s_2 \mid \langle s, s_1 \rangle \in \llbracket \phi \rrbracket, \langle s_2, s' \rangle \in \llbracket \psi \rrbracket\})
\]

\[
= \mu(\{s_1 \mid \langle s, s_1 \rangle \in \llbracket \phi \rrbracket, \langle s, s' \rangle \in \llbracket \psi \rrbracket\}).
\]

Consequently,

\[
\langle s, s' \rangle \models \phi \circ_1 \psi \iff \langle s, s_1 \rangle \models \phi \land \langle s_1, s' \rangle \models \psi \text{ for \( \mu \)-almost all } s_1.
\]

(here \( \mu \)-almost all \( s_1 \) means as usual that the set of all \( s_1 \) for which the property does not hold has \( \mu \)-measure 0). More generally,

\[
\langle s, s' \rangle \models \phi \circ_q \psi \iff \langle s, s_1 \rangle \models \phi \land \langle s_1, s' \rangle \models \psi \text{ for all } s_1 \text{ from a Borel set } S_0 \text{ with } \mu(S_0) \geq q.
\]

Finally, put \( K_\otimes(s, s') := \delta_{(s,s')} \), then \( \langle s, s' \rangle \models \otimes_q \psi \iff \langle s', s \rangle \models \phi \), for all rational \( q \) with \( 0 \leq q \leq 1 \) (which is evidently independent of \( q \)), and let

\[
K_1(s, s') := \begin{cases} 0, & s \neq s' \\ \delta_{(s,s)}, & s = s' \end{cases}
\]
(here 0 is the null measure), then
\[ \langle s, s' \rangle \models 1 \iff s = s'. \]

Note that in general we did exclude modal constants, i.e., modal operators of arity 0, when defining modal similarity types. The example shows that it is possible to include them nevertheless. ♦

6.1.2 Refinements

Given a nondeterministic and a stochastic interpretation, we want to compare both. Intuitively, the stochastic interpretation is more precise than its nondeterministic cousin: whereas nondeterministically we can only talk about possibilities, we can assign weights to these possibilities using probabilities. To say that after a certain input the output will be \(a\), \(b\) or \(c\) conveys certainly less information than saying that the probabilities for these outputs will be, respectively, \(p(a) = 1/100\), \(p(b) = 1/50\) and \(p(c) = 97/100\).

Since negation has its own problems, we will restrict the discussion to the negation free logic \(\text{Mod}_1(\tau, P)\), and we will deal in the present Section 6.1.2 exclusively with stochastic relations which assign always the whole space probability one.

**Definition 6.1.7** Let \(\mathcal{R}\) and \(\mathcal{K}\) be a nondeterministic and a stochastic \(\tau\)-Kripke model, and assume that \(K_\Delta(s)(S \times \ldots \times S^{\rho(\Delta)}) = 1\) holds for each \(s \in S\) (we will call these models probabilistic). \(\mathcal{K}\) is said to refine \(\mathcal{R}\) (abbreviated as \(\mathcal{K} \supseteq \mathcal{R}\)) iff \([\varphi]_\mathcal{K} \subseteq [\varphi]_\mathcal{R}\) holds for all \(\varphi \in \text{Mod}_1(\tau, P)\).

Consequently, given the interpretations \(\mathcal{K}\) and \(\mathcal{R}\), we have \(\mathcal{K} \supseteq \mathcal{R}\) if \(\mathcal{R}, s \models \varphi\) holds only if \(\mathcal{K}, s \models \varphi\) is true for each formula \(\varphi\) in the negation free fragment of the logic.

We will investigate the relationship between nondeterministic and stochastic satisfaction by showing that for each stochastic interpretation \(\mathcal{K}\) we can find a nondeterministic one \(\mathcal{R}\) with \(\mathcal{K} \supseteq \mathcal{R}\) by simply taking all possible state changes and making it into a Kripke model. Conversely, we will look into the possibility of refining a given nondeterministic Kripke model into a stochastic model. This requires some topological assumptions (for otherwise the notion all possible states cannot be made precise). Thus from now on the state space \(S\) is a Polish space with its Borel sets as \(\sigma\)-algebra.

The set of all states possible for a probability \(\mu\) on a Polish space is captured through the support of a probability \(\mu\) (see [73, Theorem II.2.1]).

**Lemma 6.1.8** Let \(\mu \in \mathfrak{S}(X), \mu \neq 0\) for the Polish space \(X\). Then there exists a unique closed set \(C_\mu \subseteq X\) with the following properties

1. \(\mu(C_\mu) = \mu(X)\),

2. if \(D \subseteq X\) is a closed set with \(\mu(D) = \mu(X)\), then \(C_\mu \subseteq D\),

3. \(x \in C_\mu\) iff \(\mu(U) > 0\) for all open neighborhoods \(U\) of \(x\).

**Proof** 1. Let
\[ U_\mu := \bigcup \{U \mid U\ \text{open with } \mu(U) = 0\}, \]
then $U_\mu$ is open. There exist countably many open sets $(U_n)_{n \in \mathbb{N}}$ with $U_\mu = \bigcup_{n \in \mathbb{N}} U_n$. This is so since $X$ is Polish, hence has a countable base for its topology. Thus $\mu(U_\mu) \leq \sum_{n \in \mathbb{N}} \mu(U_n) = 0$. Put $C_\mu := X \setminus U_\mu$, then plainly $C_\mu$ is closed with $\mu(C_\mu) = \mu(X)$. If $D \subseteq X$ is closed with $\mu(D) = \mu(X)$, the construction of $U_\mu$ yields $X \setminus D \subseteq U_\mu$, thus $C_\mu \subseteq D$.

2. If $x \in C_\mu$ and $U$ is an open neighborhood of $x$ with $\mu(U) = 0$, we have $U \subseteq U_\mu$, which contradicts $x \in C_\mu$. If $x \notin C_\mu$, the set $U_\mu$ is an open neighborhood if $x$ with $\mu(U_\mu) = 0$. 

**Definition 6.1.9** Let $0 \neq \mu \in \mathcal{S}(X)$ for the Polish space $X$. The set $C_\mu$ constructed in Lemma 6.1.8 is denoted by $\text{supp}(\mu)$ and is called the support of $\mu$.

Having a look at the properties of the support in Lemma 6.1.8, we see that this is exactly what we want: a set in which all points have the property that all neighborhoods have positive measure.

**Proposition 6.1.10** Let $\mathcal{K} = (S, (K_\Delta)_{\Delta \in \mathcal{O}}, V)$ be a probabilistic $\tau$-Kripke model. Define for the modal operator $\Delta \in \mathcal{O}$ the set valued map

$$R^K_\Delta(s) := \text{supp}(K_\Delta(s)).$$

Put

$$R^K = (S, (R^K_\Delta)_{\Delta \in \mathcal{O}}, V),$$

then $\mathcal{K} \otimes R^K$.

**Proof** The proof proceeds by induction on the structure of the formulas. Assume that $\Delta$ is a modal operator, and that we know $[\varphi_i]_{R^K} \subseteq [\varphi_i]_{R^K}$ for $1 \leq i \leq \rho(\Delta)$. Now suppose $R^K_\Delta(s) \neq \Delta_1(\varphi_1, \ldots, \varphi_{\rho(\Delta)})$ for some state $s$. Thus $R^K_\Delta(s) \cap [\varphi_1]_{R^K} \times \ldots \times [\varphi_{\rho(\Delta)}]_{R^K} = \emptyset$, and, consequently, by the hypothesis, $R^K_\Delta(s) \cap [\varphi_1]_{R^K} \times \ldots \times [\varphi_{\rho(\Delta)}]_{R^K} = \emptyset$. But this means $K_\Delta(s)([\varphi_1]_{R^K} \times \ldots \times [\varphi_{\rho(\Delta)}]_{R^K}) < 1$, hence $\mathcal{K}, s \not\models \Delta_1(\varphi_1, \ldots, \varphi_{\rho(\Delta)})$. 

Thus each probabilistic Kripke model carries a nondeterministic one with it, and it refines this companion (one is tempted to perceive this as a nondeterministic shadow: a shadow as a coarser, black-and-white image of a probably more colorful and picturesque original).

It will be shown now that the converse of Proposition 6.1.10 is also true: Given a nondeterministic Kripke model, there exists a stochastic one refining it. Intuitively, and in the finite case, one simply assigns a uniform weight as a probability to all possible outcomes; this is actually the starting point for the non-standard approach to probability, see [60, Example II.2.1].

This is basically what we will do here, too, but we have to be a bit more careful since in an uncountable setting this idea requires some additional underpinning. It is provided by the structure of the support map when combined with a stochastic relation, yielding a set-valued map with favorable properties.

It is immediate that the support produces a measurable relation for a probabilistic relation $K : Y \leadsto Z$: put

$$R_K := \{\langle y, z \rangle \in Y \times Z | z \in \text{supp}(K(y))\},$$

then

$$(\forall R_K)(F) = \{y \in Y | K(y)(F) = 1\}$$
for the closed set $F \subseteq Z$, and

$$(\exists R_K)(G) = \{ y \in Y | K(y)(G) > 0 \}$$

for the open set $G \subseteq Z$. Both sets are measurable.

It is also plain that a representation of $R$ through a stochastic relation $K$ which is given by

$$(\ast) \forall y \in Y : R(y) = \supp(K(y))$$

implies that $R$ has to be a measurable relation.

Given a set-valued relation $R$, a probabilistic relation $K$ with $(\ast)$ can be found. For this, $R$ has to take closed values, and a condition of measurability is imposed. We will obtain the existence of such a relation from Proposition A.2.7.

**Lemma 6.1.11** Let $R \subseteq Y \times Z$ be a measurable relation for $Y, Z$ Polish. There exists a probabilistic relation $K : Y \rightsquigarrow Z$ such that $R(y) = \supp(K(y))$ holds for each $y \in Y$.

**Proof** Because $R$ is measurable we obtain from Proposition A.2.7, part 1, a sequence $(f_n)_{n \in \mathbb{N}}$ of measurable maps $f_n : Y \to Z$ such that $\{f_n(y) \mid n \in \mathbb{N}\}$ is dense in $R(y)$ for each $y \in Y$. Define $K_n(y) := \sum_{i=1}^{\infty} 2^{-i} \cdot \delta_{f_i(y)}$, then $K_n : Y \rightsquigarrow Z$, with

$\supp(K_n(y)) = \{ f_i(y) \mid 1 \leq i \leq n \} \subseteq R(y)$.

It is not difficult to see that

$K_n(y) \to_w K(y) := \sum_{j \in \mathbb{N}} 2^{-j} \cdot \delta_{f_j(y)}$,

that $K : Y \rightsquigarrow Z$, and that $\supp(K(y)) = \text{cl}(\{f_j(y) \mid j \in \mathbb{N}\}) = R(y)$. \[ \]

Thus we can find a probabilistic Kripke structure refining a given nondeterministic one, provided we impose a measurability condition:

**Proposition 6.1.12** Suppose $\mathcal{R} := (S, (R_\Delta)_{\Delta \in \mathcal{O}}, V)$ is a nondeterministic $\tau$-Kripke model such that

1. $V(p) \in \mathcal{B}(S)$ for all $p \in P$,
2. $R_\Delta$ is a measurable relation on $S \times S^{\rho(\Delta)}$ for each $\Delta \in \mathcal{O}$.

Then there exists a probabilistic $\tau$-Kripke model $\mathcal{K} = (S, (K_\Delta)_{\Delta \in \mathcal{O}}, V)$ with $K \otimes \mathcal{R}$.

**Proof** Applying Lemma 6.1.11, find for each modal operator $\Delta \in \mathcal{O}$ a transition probability $K_\Delta : S \rightsquigarrow S^{\rho(\Delta)}$ such that $R_\Delta(s) = \supp(K_\Delta(s))$ holds for all $s \in S$. The argumentation in the proof of Lemma 6.1.10 establishes the claim. \[ \]
Corollary 6.1.13 Let \( R \) be a nondeterministic \( \tau \)-Kripke model satisfying the conditions of Proposition 6.1.12. Assume that \( K_i := (S_i, (K_{\Delta,i})_{\Delta \in O}, V_i) \) is a probabilistic \( \tau \)-Kripke model with \( K_i \circlearrowleft R \) for each \( i \in \mathbb{N} \). Let \( (\alpha_i)_{i \in \mathbb{N}} \) be a sequence of positive real numbers such that \( \sum_{i \in \mathbb{N}} \alpha_i = 1 \), and define for \( \Delta \in O \) the stochastic relation

\[
K_\Delta(s) := \sum_{i \in \mathbb{N}} \alpha_i \cdot K_{\Delta,i}(s).
\]

Then \( (S, (K_\Delta)_{\Delta \in O}, V) \circlearrowleft R \).

Proof Let \( (\mu_i)_{i \in \mathbb{N}} \) be a sequence of probability measures. Since all \( \alpha_i \) are positive, the definition of the support function yields that

\[
\text{supp}(\sum_{i \in \mathbb{N}} \alpha_i \cdot \mu_i) = \text{cl} \left( \bigcup_{i \in \mathbb{N}} \text{supp}(\mu_i) \right)
\]

holds. Thus \( R_\Delta \) equals \( \text{supp}(K_\Delta) \). The assertion now follows from Proposition 6.1.12.

A substantial generalization in which replacing the infinite convex combination is replaced by a suitable integration is proposed in [26]. Thus we know that not only a probabilistic \( \tau \)-Kripke model is the refinement of a probabilistic one, but also that refinements offer a considerable degree of freedom, because they are closed under countable convex combinations (in fact, it can also be shown that it is closed under integration as the generalization of convex combinations). This supports the intuitive feeling that a probabilistic model conveys much more information than a nondeterministic one, but that it is also much harder to obtain.

6.1.3 Bisimulations for Kripke Models

This section investigates morphisms for stochastic \( \tau \)-Kripke models; we want to know whether bisimilarity and mutually identical theories are equivalent also for this general case. To this end we first discuss morphisms that are based on morphisms for stochastic relations (a \( \tau \)-Kripke model is built from a family of stochastic relations, after all), and indicate that this notion of morphism is not adequate for our purposes and propose the notion of a strong morphism. We show that strong morphisms are suitable for our purposes.

Assume first that the set \( P \) of propositional letters is empty, rendering the initial discussion a bit less technical. Then a stochastic \( \tau \)-Kripke model \( K := (S, (K_\Delta)_{\Delta \in O}) \) is determined through the Polish state space \( S \) and the family \( K_\Delta : S \rightsquigarrow S^{\rho(\Delta)} \) of stochastic relations. A morphism

\[
\Phi : (S, (K_\Delta)_{\Delta \in O}) \rightarrow (S', (K'_\Delta)_{\Delta \in O})
\]

for stochastic \( \tau \)-Kripke models is then a family \( \Phi = ((\phi_\Delta, \psi_\Delta)_{\Delta \in O}) \) of morphisms

\[
(\phi_\Delta, \psi_\Delta) : (S, S^{\rho(\Delta)}, K_\Delta) \rightarrow (S', (S')^{\rho(\Delta)}, K'_\Delta)
\]

for the associated relations.
Consider a modal operator $\triangle$. The $\sigma$-algebra $\mathcal{A}_\triangle$ generated by
\[
\{[\varphi_1]_K \times \ldots \times [\varphi_{n(\triangle)}]_K \mid \varphi_1, \ldots, \varphi_{n(\triangle)} \in \text{Mod}_\triangle(\tau, P)\}
\]
is evidently countably generated, thus gives rise to a smooth equivalence relation $\beta_\triangle$ on $S_{n(\triangle)}$, and the relation
\[
s_{\alpha_\triangle}s' \iff \forall B \in \mathcal{A}_\triangle : K_\triangle(s)(B) = K_\triangle(s')(B)
\]
is smooth due to $\mathcal{A}_\triangle$ being countably generated. Consequently, $(\alpha_\triangle, \beta_\triangle)$ is a congruence for $K_\triangle : S \leadsto S_{n(\triangle)}$, and if $\mathcal{K}$ is non-degenerate, this congruence is non-trivial.

Let $\mathcal{K}' = (S', (K'_\triangle)_{\triangle \in T})$ be another $\tau$-Kripke model which is equivalent to the first one in the sense that for the states the corresponding theories mutually coincide. To be more precise:

**Definition 6.1.14** The stochastic $\tau$-Kripke models $\mathcal{K}$ and $\mathcal{K}'$ are said to be equivalent (abbreviated as $\mathcal{K} \sim \mathcal{K}'$) iff $\{\text{Th}_{\mathcal{K}}(s) \mid s \in S\} = \{\text{Th}_{\mathcal{K}'}(s') \mid s' \in S'\}$.

Thus $\mathcal{K} \sim \mathcal{K}'$ iff given $s \in S$ there exists $s' \in S'$ such that $\text{Th}_{\mathcal{K}}(s) = \text{Th}_{\mathcal{K}'}(s')$, and vice versa.

Assume both $\mathcal{K}$ and $\mathcal{K}'$ are non-degenerate. Construct for $\mathcal{K}'$ the congruence $(\alpha'_\triangle, \beta'_\triangle)$ for each modal operator $\triangle$ as above, then it can be shown that $\mathcal{K} \sim \mathcal{K}'$ implies that the congruences $(\alpha_\triangle, \beta_\triangle)$ and $(\alpha'_\triangle, \beta'_\triangle)$ are equivalent. From Proposition 5.3.3 we see that $K_\triangle$ and $K'_\triangle$ are bisimilar for each modal operator $\triangle$, so that there exists a span of morphisms
\[
(S, S_{n(\triangle)}, K_\triangle) \overset{(\phi_\triangle, \psi_\triangle)}{\longleftarrow} (A_\triangle, B_\triangle, M_\triangle) \overset{(\phi'_\triangle, \psi'_\triangle)}{\longrightarrow} (S', (S')_{n(\triangle)}, K'_\triangle).
\]

This is rather satisfying from the point of view of stochastic relations, but not when considering stochastic $\tau$-Kripke models. This is so since in general $((A_\triangle, B_\triangle, M_\triangle)_{\triangle \in T})$ fails to be such a model, because there is no way to guarantee that all $A_\triangle$ coincide with, say, a Polish space $T$, and so that $B_\triangle$ equals $T_{n(\triangle)}$.

Consequently, we have to strengthen the requirements for a morphism in order to achieve some uniformity. This will be done now, and we admit propositional letters again.

The basic idea is to have just one map $\phi$ between the state spaces so that
\[
K'_{\triangle}(\phi(s))(A) = K_\triangle(s)((\{s_1, \ldots, s_{n(\triangle)}\}|(\phi(s_1), \ldots, \phi(s_{n(\triangle)})) \in A))
\]
holds for each state $s \in S$ and each Borel set $A \subseteq (S')_{n(\triangle)}$, making the diagram
\[
\begin{array}{ccc}
S & \xrightarrow{\phi} & S' \\
\downarrow K_\triangle & & \downarrow K'_\triangle \\
\mathcal{G}(S_{n(\triangle)}) & \xrightarrow{\mathcal{G}(\phi_{n(\triangle)})} & \mathcal{G}((S')_{n(\triangle)})
\end{array}
\]
commutative (where $\phi^n : (x_1, \ldots, x_n) \mapsto (\phi(x_1), \ldots, \phi(x_n))$ distributes $\phi$ into the components), and we want to have $s \in V(p)$ iff $\phi(s) \in V'(p)$ for each propositional letter $p$. This leads to
Definition 6.1.15 Let $\mathcal{K} := (S, (K_\Delta)_{\Delta \in O}, V)$ and $\mathcal{K'} := (S', (K'_\Delta)_{\Delta \in O}, V')$ be stochastic $\tau$-Kripke models. A strong morphism $\phi : \mathcal{K} \to \mathcal{K'}$ is determined through a measurable and surjective map $\phi : S \to S'$ so that these conditions are satisfied:

1. $\forall p \in P : V(p) = \phi^{-1}[V'(p)]$,
2. for each modal operator $\Delta$,

$$K'_\Delta \circ \phi = \mathcal{G}(\phi^{\rho(\Delta)}) \circ K_\Delta$$

holds.

Thus, if $\phi : \mathcal{K} \to \mathcal{K'}$ is a strong morphism, then $$(\phi, \phi^{\rho(\Delta)}): (S, S^{\rho(\Delta)}, K_\Delta) \to (S', S'^{\rho(\Delta)}, K'_\Delta)$$

is a morphism between the corresponding stochastic relations for each modal operator $\Delta \in O$. Note that we take also the propositional letters into account.

It is clear that stochastic $\tau$-Kripke models over general measurable spaces form a category $pKripke$ with this notion of morphism, because the composition of strong morphisms is again a strong morphism, and because the identity is a strong morphism, too. Furthermore, each modal operator $\Delta$ induces a functor $F_\Delta : pKripke \to Stoch$ which forgets all but $K_\Delta$. We will below make (rather informal) use of this functor.

Because we work on the safe grounds of a category, we have bisimulations at our disposal, which can be defined again as spans of strong morphisms:

Definition 6.1.16 The stochastic $\tau$-Kripke models $\mathcal{K}_1$ and $\mathcal{K}_2$ are called strongly bisimilar iff

1. there exists a mediating stochastic $\tau$-Kripke model $\mathcal{M}$ and strong morphisms $\phi_1 : \mathcal{K}_1 \to \mathcal{M}$ and $\phi_2 : \mathcal{M} \to \mathcal{K}_2$,

2. the $\sigma$-algebra $\phi_1^{-1}[\mathcal{B}(S_1)] \cap \phi_2^{-1}[\mathcal{B}(S_2)]$ is non-trivial (here $S_i$ is the state space of $\mathcal{K}_i$, $i = 1, 2$).

Since the product $\sigma$-algebra is the smallest $\sigma$-algebra which contains all the measurable rectangles, it is not difficult to see that $\phi_1^{-1}[\mathcal{B}(S_1)] \cap \phi_2^{-1}[\mathcal{B}(S_2)]$ is non-trivial iff for each modal operator $\Delta \in O$ the $\sigma$-algebra

$$\prod_{i=1}^{\rho(\Delta)} \phi_i^{-1}[\mathcal{B}(S_1)] \cap \prod_{i=1}^{\rho(\Delta)} \phi_i^{-1}[\mathcal{B}(S_2)]$$

is non-trivial. Thus condition 2 in Definition 6.1.16 will imply that this notion of bisimilarity is compatible to the one used for stochastic relations in general.

We will show that $\mathcal{K} \sim \mathcal{K'}$ iff $\mathcal{K}$ and $\mathcal{K'}$ are strongly bisimilar, provided the models are based on Polish spaces. Fix the stochastic $\tau$-Kripke models $\mathcal{K} := (S, (K_\Delta)_{\Delta \in O}, V)$ and $\mathcal{K'} := (S', (K'_\Delta)_{\Delta \in O}, V')$.

It is well-known that morphisms preserve theories for the Hennessy-Milner logic [20]. This is also true for stochastic relations:
Lemma 6.1.17 If $\phi : K \to K'$ is a strong morphism, then $Th_{K'}(s) = Th_K(\phi(s))$ holds for all states $s \in S$.

Proof. We show by induction on the formula $\varphi \in \mathcal{M}_{\Delta}(\tau, P)$ that

$$K, s \models \varphi \iff K', \phi(s) \models \varphi$$

holds; putting it slightly different, we want to show

$$(*) [\varphi]_K = [\varphi]_{K'}$$

for all these $\varphi$.

1. The equivalence relations involved are all smooth, so it first has to be demonstrated that each pair forms indeed a congruence. Assume that

$$F$$

and

$$\triangle$$

are equivalent and non-trivial congruences on the stochastic relations

$$K \equiv K'$$

then

$$\triangle$$

is smooth, and we know that the

holds; putting it slightly different, we want to show

$$(*) [\varphi]_K = [\varphi]_{K'}$$

for all these $\varphi$.

2. If $\varphi = p \in P$, this follows from $V(p) = \phi^{-1}[V'(p)]$. The interesting case in the induction step is the application of a $n$-ary modal operator $\Delta_q$ with rational $q$. Suppose the assertion is true for $[\varphi_1]_K, \ldots, [\varphi_n]_K$, then

$$K, s \models \Delta_q(\varphi_1, \ldots, \varphi_n) \iff K_{\Delta}(s)([\varphi_1]_K \times \ldots \times [\varphi_n]_K) \geq q$$

$$\frac{(\dagger)}{\dagger} K_{\Delta}(s)([\varphi_1]_K \times \ldots \times [\varphi_n]_K) \geq q$$

$$\iff (\Theta(\phi^n) \circ K_{\Delta})(s)([\varphi_1]_K \times \ldots \times [\varphi_n]_K) \geq q$$

$$\frac{(\dagger)}{\dagger} K_{\Delta}(\phi(s))( [\varphi_1]_{K'} \times \ldots \times [\varphi_n]_{K'}) \geq q$$

$$\iff K', \phi(s) \models \Delta_q(\varphi_1, \ldots, \varphi_n).$$

In (\dagger) we use reformulation (*) for the induction hypothesis, in (\dagger) we make use of the defining equation of a (strong) morphism. \dagger

Define the equivalence relation $\alpha$ on state space $S$ through

$$s_1 \alpha s_2 \iff Th_K(s_1) = Th_K(s_2),$$

thus two states are $\alpha$-equivalent iff they satisfy exactly the same formulas in $\mathcal{M}_{\Delta}(\tau, P)$; in a similar way $\alpha'$ is defined on $S'$. Because we have at most countably many formulas, $\alpha$ and $\alpha'$ are smooth equivalence relations. Define the equivalence relation $\beta_{\Delta}$ on $S^{\rho(\Delta)}$ through

$$\langle s_1, \ldots, s_{\rho(\Delta)} \rangle \beta_{\Delta} \langle t_1, \ldots, t_{\rho(\Delta)} \rangle \iff s_1 \alpha t_1 \wedge \ldots \wedge s_{\rho(\Delta)} \alpha t_{\rho(\Delta)},$$

then $\beta_{\Delta}$ is smooth, and we know that the $\sigma$-algebra of $\beta$-invariant sets can be written in terms of the $\alpha$-invariant sets, viz., $\mathcal{I}NV(B(S^{\rho(\Delta)}), \beta_{\Delta}) = \bigotimes_{i=1}^{\rho(\Delta)} \mathcal{I}NV(B(S), \alpha)$ (see Lemma 5.1.18). The relation $\beta_{\Delta}'$ is defined in the same way for $\alpha'$. The equivalence of $K$ and $K'$ makes these relations into equivalent congruences:

Lemma 6.1.18 If $K \sim K'$ for the non-degenerate Kripke models $K$ and $K'$, then $\langle \alpha, \beta_{\Delta} \rangle$ and $\langle \alpha', \beta_{\Delta}' \rangle$ are equivalent and non-trivial congruences on the stochastic relations $F_{\Delta}(K)$ and $F_{\Delta}(K')$.

Proof. The equivalence relations involved are all smooth, so it first has to be demonstrated that each pair forms indeed a congruence. Assume that $s_1 \alpha s_2$ holds, then

$$K_{\Delta}(s_1)([\varphi_1]_K \times \ldots \times [\varphi_{\rho(\Delta)}]_K) = K_{\Delta}(s_2)([\varphi_1]_K \times \ldots \times [\varphi_{\rho(\Delta)}]_K)$$

186
forms a generator for $INV(B(S^{\rho(\Delta)}), \beta_\Delta)$, we see that $(\alpha, \beta_\Delta)$ is a congruence for $F_{\Delta}(K)$. The same arguments show that also $(\alpha', \beta_{\Delta}')$ is a congruence for $F_{\Delta}(K')$.

2. $A_0 := \{[\varphi]_K | \varphi \in \mathcal{M}_\alpha(\tau, P)\}$ is a countable generator of the $\sigma$-algebra $INV(B(S), \alpha)$, and since the logic is closed under conjunction, $A_0$ is closed under finite intersections. Given $s \in S$ there exists $s' \in S'$ such that $Th_{K'}(s) = Th_{K'}(s')$ holds; define $\Upsilon([s]_\alpha) := [s']_{\alpha'}$, then $\Upsilon : S/\alpha \to S'/\alpha'$ is well defined, and $\Upsilon([s]_K) = [\varphi]_{K'}$ holds. Consequently, $\{\Upsilon_A | A \in A_0\}$ generates $INV(B(S'), \alpha')$, and the construction implies that

$$\bigcap\{\Upsilon_A | s \in A \in A_0\} \cap \bigcap\{S' \setminus \Upsilon_A | s \notin A \in A_0\} = [s']_{\alpha'}.$$

Hence $\alpha$ spawns $\alpha'$ via $(\Upsilon, A_0)$.

3. The construction of $\beta_\Delta$ yields

$$[\{s_1, \ldots, s_{\rho(\Delta)}\}]_{\beta_\Delta} = [s_1]_{\alpha} \times \ldots \times [s_{\rho(\Delta)}]_{\alpha}.$$

An argument very similar to that used above shows that $\beta_\Delta$ spawns $\beta_{\Delta}'$ via $(\Theta, B_0)$, where

$$\Theta : [\{s_1, \ldots, s_{\rho(\Delta)}\}]_{\beta_\Delta} \mapsto \Upsilon([s_1]_\alpha) \times \ldots \times \Upsilon([s_{\rho(\Delta)}]_\alpha),$$

and $B_0$ is defined above.

4. An argumentation very close to the first part of the proof shows that $Th_{K}(s) = Th_{K'}(s')$ for $s \in S, s' \in S'$ implies for all formulas $\varphi_1, \ldots, \varphi_{\rho(\Delta)}$ that

$$K_{\Delta}(s)([\varphi_1]_K \times \ldots \times [\varphi_{\rho(\Delta)}]_K) = K'_{\Delta}(s')(\varphi_1]_{K'} \times \ldots \times [\varphi_{\rho(\Delta)}]_{K'})$$

(ep. part 2 of the proof of Lemma 6.1.17). Thus $(\alpha, \beta_\Delta) \varpropto (\alpha', \beta_{\Delta}')$, and in the same way, interchanging the roles of $K$ and $K'$, we infer $(\alpha', \beta_{\Delta}') \varpropto (\alpha, \beta_\Delta)$.

5. Because $K$ is non-degenerate, the $\sigma$-algebra

$$INV(B(S), \alpha) = \sigma([\{\varphi\}_K | \varphi \in \mathcal{M}_\alpha(\tau, P))]$$

is non-trivial, because

$$INV(B(S^{\rho(\Delta)}), \beta_\Delta) = \bigotimes_{i=1}^{\rho(\Delta)} INV(B(S), \alpha),$$

we see that $INV(B(S^{\rho(\Delta)}), \beta_\Delta)$ contains a set of the form $B^{\rho(\Delta)}$ for some $B$ with $\emptyset \neq B \neq S$, thus we may conclude that $\beta_\Delta$ is not the universal relation. Thus $(\alpha, \beta_\Delta)$ is a non-trivial congruence. Replacing $K$ by $K'$, this is also established for the congruence $(\alpha', \beta_{\Delta}')$. This completes the proof. $\square$

Accordingly, we know from Proposition 5.3.3 that for equivalent Kripke models $K$ and $K'$ and for each modal operator $\Delta$ the stochastic relations $F_{\Delta}(K)$ and $F_{\Delta}(K')$ are bisimilar. All the mediating relations can be collected to form a mediating Kripke model. This requires, however, that we know a wee bit about the internal structure of the semi-pullback which is constructed along the way. We will see this in the proof of the following result, the Hennessy-Milner Theorem for stochastic $\tau$-Kripke models:
**Theorem 6.1.19** Assume that $\mathcal{K}$ and $\mathcal{K}'$ are non-degenerate stochastic $\tau$-Kripke models over Polish spaces, then the following statements are equivalent:

1. $\mathcal{K}$ and $\mathcal{K}'$ are strongly bisimilar;

2. $\mathcal{K} \sim \mathcal{K}'$.

**Proof** 1. Because $1 \Rightarrow 2$ follows from Lemma 6.1.17, we may concentrate on the proof for $2 \Rightarrow 1$.

2. Since $\mathcal{K} \sim \mathcal{K}'$, we know from Lemma 6.1.18 that the congruences $(\alpha, \beta_\Delta)$ and $(\alpha', \beta'_\Delta)$ are equivalent for each modal operator $\Delta$. Let $M_\Delta = (M_\Delta, N_\Delta, L_\Delta)$ be the mediating stochastic relation, which exists by Proposition 5.3.3. The proof of Theorem 4.3.2 shows that $(n := \rho(\Delta))$

$$M_\Delta = \{ (s, s') \in S \times S' \mid s (\alpha + \alpha') s' \},$$

$$N_\Delta = \{ (s_1, s_1', \ldots, s_n, s_n') \in (S \times S')^n \mid s_i (\alpha + \alpha') s'_i \text{ for } 1 \leq i \leq n \},$$

which may be rendered Polish spaces. Note that $S'' := M_\Delta$ does not depend at all on the modal operator, and that $N_\Delta$ depends only on its arity. Furthermore, we may infer for the $P$ – Stoch-morphisms

$$F_\Delta(\mathcal{K}) \xrightarrow{f_\Delta} M_\Delta \xrightarrow{f'_\Delta} F_\Delta(\mathcal{K}')$$

that $f_\Delta = (\pi_{1,S}, \pi_{1,S}'), f'_\Delta = (\pi_{2,S'}, \pi_{2,S'})$ holds, where the $\pi$ denote the projections.

Now define for the propositional letter $p \in P$

$$W(p) := \{ (s, s') \in M_\Delta \mid s \in V(p), s' \in V'(p) \},$$

then it is immediate that the equations $W(p) = \pi_1^{-1}[V(p)] = \pi_2^{-1}[V'(p)]$ hold. Consequently, $M := (S'', (L_\Delta)_{\Delta \in \mathcal{O}}, W)$ is a stochastic $\tau$-Kripke model with

$$\mathcal{K} \xrightarrow{\pi_{1,S}} M \xrightarrow{\pi_{2,S'}} \mathcal{K}'$$

in pKripke.

3. We need finally to show that the $\sigma$-algebra $\pi_{1,S}^{-1}[B(S)] \cap \pi_{2,S'}^{-1}[B(S')]$ is non-trivial. This is essentially the same argument as the one used in the third part of the proof to Proposition 5.3.3. Since $\mathcal{K}$ is non-degenerate, we can find a formula $\phi$ with $\emptyset \neq [\phi]_\mathcal{K} \neq S$. The set $[\phi]_\mathcal{K}$ is an $\alpha$-invariant Borel subset of $S$. We know from the proof of Lemma 6.1.18 that $\alpha$ spawns $\alpha'$ via $(\Upsilon, \{ [\phi]_\mathcal{K} \mid \phi \in \mathcal{M}(\pi, P) \})$ for some suitably chosen $\Upsilon$. Thus $\pi_{1,S}^{-1}[\phi]_\mathcal{K} = \pi_{2,S'}^{-1}[\Upsilon[\phi]_\mathcal{K}]$, consequently we see that $\pi_{1,S}^{-1}[\phi]_\mathcal{K} \in \pi_{1,S}^{-1}[B(S)] \cap \pi_{2,S'}^{-1}[B(S')]$. Since $\emptyset \neq [\phi]_\mathcal{K} \neq S$ we conclude that $\emptyset \neq \pi_{1,S}^{-1}[\phi]_\mathcal{K} \neq M_\Delta$, hence the $\sigma$-algebra in question is indeed not trivial. $\dagger$

Looking back at the development, it may be noted that Theorem 6.1.19 is derived from Proposition 5.3.3, hence from a condition that arose from the consideration of stochastic relations alone. This is in marked contrast to the proofs proposed in [20, 25] which start from the logic and develop the properties of equivalent congruences implicitly.

The following example gives a brief illumination.
Example 6.1.20 Consider Kripke models for the basic modal language that has just one modal operator, traditionally denoted by $\Diamond$, which is unary. Assume that there are at least two propositional letters. Let $\mathcal{K} = (S, K_\Diamond, V)$ and $\mathcal{L} = (S, L_\Diamond, W)$ be stochastic Kripke models such that $\{ (V(p), W(p)) \mid p \in P \}$ is a block for $K_\Diamond, L_\Diamond$ (see Definition 5.3.5). Then $\mathcal{K} \sim \mathcal{L}$.

This is so since the Kripke models are strongly bisimilar by Corollary 5.3.6 and by Theorem 6.1.19. ♦

6.2 Infinite Paths for Interpreting Temporal Logics

The interpretation of modal logics rested on relational properties, e.g. we said that

$$s \models \triangle q(\varphi_1, \ldots, \varphi_{\rho(\Delta)}) \iff K_\Delta(s) ([\varphi_1] \times [\varphi_{\rho(\Delta)}]) \geq q,$$

so we used the relation associated with the modal operator $\Delta$ to associate a probability to the finite path of those words of length $\rho(\Delta)$ that satisfy $\varphi_1, \ldots, \varphi_{\rho(\Delta)}$. Changing to infinite paths, we basically could do the same: assign a probability to that set of paths that we want to consider. This requires relations that work on those paths, but they are usually not given offhand. Modelling a reactive system with possibly non-terminating computations, we have to piece together the probabilities for these infinite paths from their finite components. They are usually given through relations that describe what happens in a single step. This is what we will discuss next.

Since finding an adequate probability is sometimes a bit intricate, we will first have a look at the measure-theoretic mechanisms. Then we will apply this by showing how the machinery developed so far may be put to use for discussing $\text{pCTL}^*$, a fairly popular logic for model checking. We will discuss in depth some properties of the continuous time stochastic logic $\text{CSL}$, which incorporates time explicitly. It offers not only a very powerful approach to describing systems, but also an additional challenge for the treatment of its probabilistic properties, in particular to the important problem of bisimilar states.

6.2.1 Probabilities for Paths

Assume that we are given a Polish space $X$ for modelling the state space, and that we have for each step $n$ a stochastic relation $J_n : X \sim X$ with $J_n(x)(X) = 1$ for all $x \in X$. We want to describe the probability for starting in a state $x$ and walking the infinite path $\langle x_1, x_2, \ldots \rangle$. Since we are usually not able to assign probabilities to individual points, we ask for the probability $J(x)(A)$ that we will run through an infinite path that is an element of the Borel set $A \subseteq X^\infty$. The easy way out of this question would be to define $J(x)$ as the infinite product $\bigotimes_{n \in \mathbb{N}} J_n(x)$, so if $A$ happens to have the shape $A = (A_1 \times \cdots \times A_k) \times X \times X \times \ldots$, we would have

$$J(x)(A_1 \times \cdots \times A_k \times X \times X \times \ldots) = J_1(x)(A_1) \cdots J_k(x)(A_k).$$

But this is in general not really adequate, since this would suggest that the steps are mutually independent, which in general they are not. One would take at step $n + 1$ at least the knowledge from step $n$ into account. Hence we have to find a finer description.
which would include the case just discussed, but which would also leave other possibilities.
Suppose we describe the probabilities for paths of length \( k \) through stochastic relations
\( J^{(k)} : X \rightsquigarrow X^k \) which are somehow based on \( (J_n)_{n \in \mathbb{N}} \), then we would want to have
\[
J^{(k+1)}(x)(A \times X) = J^{(k)}(x)(A),
\]
whenever \( A \subseteq X^k \) is a Borel set. This is clear: after \( k \) steps we will want to turn somewhere, taking the history into account. The above property translates into
\[
J^{(k)}(x) = \mathcal{P}(\pi_{k+1})(J^{(k+1)}(x)),
\]
where \( \pi_{k+1} : X^{k+1} \rightarrow X^k \) denotes the projection \( \langle x_1, \ldots, x_{k+1} \rangle \mapsto \langle x_1, \ldots, x_k \rangle \). Thus the sequence \( (J^{(k)}(x))_{k \in \mathbb{N}} \) is a projective system of probability measures for each state \( x \in X \), see section A.3.3, where a definition of \( J^{(k)} : X \rightsquigarrow X^k \) is given. Consequently there are more possibilities to produce a projective system from the sequence \( (J_n)_{n \in \mathbb{N}} \) than to dwell upon the product only, as we pretended above. This will be applied now to the interpretation of two temporal logics.

### 6.2.2 pCTL*

A state formula \( \varphi \) in pCTL* is given through the syntax
\[
\varphi ::= a \mid T \mid \varphi_1 \land \varphi_2 \mid \neg \varphi \mid [\psi_1 \sqsubseteq_\mathcal{C} \psi_2] \sqsupseteq q,
\]
where \( \psi \) is given through
\[
\psi ::= a \mid T \mid \psi_1 \land \psi_2.
\]
Here a countable set \( P \) of atomic formulas is given, \( a \in P, c \in \mathbb{N} \cup \{0\}, q \in \mathbb{Q} \cap [0,1] \) and \( \sim, \sqsupseteq \) are relational symbols from \( \{<, >, \leq, \geq\} \).
A state formula \([\psi_1 \sqsubseteq_\mathcal{C} \psi_2] \sqsupseteq q\) can only be used on the top level, these formulas cannot be nested. Various modal operators may be obtained as special cases, e.g. \( \Diamond_q \psi \) as \([\top \sqsubseteq_\mathcal{C} \psi] \leq q\).
Validity for a state formula which does not contain the until operator is given in the usual way. This requires an interpretation of atomic formulas; we assume that we have a map \( V : P \rightarrow B(S) \) at our disposal, where \( S \) is the Polish state space over which we interpret the logic. Thus we say for \( a \in P \) that \( s \models a \) iff \( s \in V(a) \), provided \( s \in S \) is a state. If \( \sigma \in S^\infty \) is an infinite path, we say that \( \sigma \models a \) iff \( \sigma_1 \in V(a) \), where \( \sigma_1 \) is the first component of the infinite path \( \sigma \). Moreover
\[
\sigma \models [\psi_1 \sqsubseteq_\mathcal{C} \psi_2] \geq 1 \iff \exists x : (\sigma_x \models \psi_2 \land \forall y \in [0,x] : \sigma_y \models \psi_1)
\]
(with \( \sigma_k \) as the \( k \)th component of \( \sigma \)).
For modeling validity of the \( \sqsubseteq \)-operator on paths, we assume that we are given a stochastic relation \( K : S \rightsquigarrow S \). We construct from \( K \) the projective system \( K^{(n)} : S \rightsquigarrow S^n \), and let \( K^\infty : S \rightsquigarrow S^\infty \) be the stochastic relation given by the projective limit of \( (K^{(n)})_{n \in \mathbb{N}} \). Then we set
\[
s \models [\psi_1 \sqsubseteq_\mathcal{C} \psi_2] \sqsupseteq q \iff K^\infty(s) (\{\sigma \in S^\infty \mid \sigma_1 = s \land \sigma \models [\psi_1 \sqsubseteq_\mathcal{C} \psi_2] \geq 1\}) \sqsupseteq q.
\]
In this way satisfaction of state formulas through the stochastic system generated from the relation $K$ can be modeled.

Denote by $[\varphi]$ and by $[\psi]$ the set of states and paths for which the state formula $\varphi$ resp. the path formula $\psi$ holds.

**Lemma 6.2.1** If $\varphi$ is a state formula, then $[\varphi] \in \mathcal{B}(S)$, if $\psi$ is a path formula, then $[\psi] \in \mathcal{B}(S^\infty)$. \(\square\)

This example demonstrates the use of the projective limit as a stochastic relation, specifically relating states to infinite sequences of states in a manner compatible with the transitions being undertaken in each step. Projective limits will also be a key player in the discussion on CSL.

### 6.2.3 CSL

The continuous time stochastic logic CSL will be analyzed in greater detail, in particular we will look at equivalence relations that are given through subsets of formulas.

Fix $P$ as a countable set of atomic propositions. We define recursively state formulas and path formulas for CSL:

**State formulas** are defined through the syntax

$$\varphi ::= \top | a | \neg \varphi | \varphi \land \varphi' | S_{\varphi_{p}}(\varphi) | P_{\varphi_{p}}(\psi).$$

Here $a \in P$ is an atomic proposition, $\psi$ is a path formula, $\ltimes$ is one of the relational operators $<, \leq, \geq, >$, and $p \in [0, 1]$ is a rational number.

**Path formulas** are defined through

$$\psi ::= X^I \varphi | \varphi U^I \varphi'$$

with $\varphi, \varphi'$ as state formulas, $I \subseteq \mathbb{R}_+$ a closed interval of the real numbers with rational bounds (including $I = \mathbb{R}_+$).

We denote the set of all state formulas by $\mathcal{L}_P$.

The operator $S_{\varphi_{p}}(\varphi)$ gives the steady-state probability for $\varphi$ to hold with the boundary condition $\ltimes p$; the formula $P$ replaces quantification: the path-quantifier formula $P_{\varphi_{p}}(\psi)$ holds for a state $s$ iff the probability of all paths starting in $x$ and satisfying $\psi$ is specified by $\ltimes p$. Thus $\psi$ holds on all paths starting from $s$ iff $s$ satisfies $P_{\varphi_{p}}(\psi)$, a path being an alternating infinite sequence $\sigma = (s_0, t_0, s_1, t_1, \ldots)$ of states $x_i$ and of times $t_i$. Note that the time is being made explicit here. The next-operator $X^I \varphi$ is assumed to hold on path $\sigma$ iff $s_1$ satisfies $\varphi$, and $t_0 \in I$ holds. Finally, the until-operator $\varphi_1 U^I \varphi_2$ holds on path $\sigma$ iff we can find a point in time $t \in I$ such that the state $\sigma @ t$ which $\sigma$ occupies at time $t$ satisfies $\varphi_2$, and for all times $t'$ before that, $\sigma @ t'$ satisfies $\varphi_1$.

The basic operators are introduced now more formally. We will also have a look at issues of measurability: the basic operators will be shown to represent measurable functions. This will help in establishing that many of the sets of paths and, derived from them, sets of states that occur in conjunction with set and path formulas are measurable, thus lie in the domain of the probabilities that we will be working with (suppose in the contrary
Setting the Stage

Fix a Polish state space $S$ over which the logic will be interpreted. A path $\sigma$ is an element of the set ${\text{PATHS}} := (S \times \mathbb{R}_+)^\infty$. Path $\sigma = (s_0, t_0, s_1, t_1, \ldots)$ may be written as $s_0 \xrightarrow{t_0} s_1 \xrightarrow{t_1} \ldots$ with the interpretation that $t_i$ is the time spent in state $s_i$. Given $i \in \mathbb{N}$, denote $s_i$ by $\sigma[i]$ as the $(i+1)$st state of $\sigma$, and let $\delta(x,i) := t_i$. Let for $t \in \mathbb{R}_+$ the index $i$ be the smallest index $k$ such that $t < \sum_{i=0}^k t_i$, and put $\sigma@t := \sigma[i]$, if $i$ is defined; set $\sigma@t := \#$, otherwise (here $\#$ is a new symbol not in $S \cup \mathbb{R}_+$). $S_{\#}$ denotes $S \cup \{\#\}$; this is a Polish space when endowed with the sum $\sigma$-algebra. The definition of $\sigma@t$ makes sure that for any time $t$ we can find a rational time $t'$ with $\sigma@t = \sigma@t'$.

We will assume that we work with a family $(K_n)_{n \in \mathbb{N}}$ of stochastic relations $K_n : S \leadsto S$, so that at the discrete time-point $n \in \mathbb{N}$ the state transitions are governed by $K_n$, hence $K_n(s)(D)$ is interpreted as the probability that the new state at $n + 1$ is a member of the Borel set $D \subseteq S$, provided the state at $n$ is $s$. In a similar way we assume that we have a family $(L_n)_{n \in \mathbb{N}}$ of stochastic relations $L_n : S \leadsto \mathbb{R}_+$ which gives the time in which transitions are triggered: suppose the system is at $n$ in state $s$, then $L_n(s)([t_1,t_2])$ is the probability that jumping from $s$ to another state occurs within the time interval $[t_1,t_2]$.

**Observation 6.2.2** The usual approach to interpreting continuous time Markov chains runs via a rate function [7]. Assume that $R$ represents the rate, then

1. $\forall s \in S : R(s)$ is a finite measure on $S$ such that $E(s) := R(s)(S) > 0$ always holds,

2. $\forall B \in \mathcal{B}(S) : s \mapsto R(s)(B)$ is a measurable function $S \rightarrow \mathbb{R}_+$.

The rate function models the transition rate: if the system is in state $s$, then the transition rate for jumping to a new state that is a member of the Borel set $D \subseteq S$ is given by $R(s)(D)$. This transition rate is assumed to be finite. We also assume in the rate model that there is no blind state, so transitions are assumed to be possible from all states, thus $E(s) > 0$.

Put

$$K_n(s)(D) := \frac{R(s)(D)}{E(s)}$$

and set for the probability of making a transition from state $s$ within $t$ time units

$$L_n(s)([0,t]) := 1 - e^{-E(s)t},$$

then

$$L_n(s)(F) = \frac{1}{E(s)} \int_F e^{-E(s)t} \, dt$$
6.2 Infinite Paths for Interpreting Temporal Logics

is independent of \( n \). Consequently, the approach discussed here fits into the usual set up to model continuous time Markov processes in the sense of [7, 22], and generalizes it. ▪

Changes of state and times of change are assumed to be stochastically independent, thus

\[ (L_n \otimes K_n)(s)(D) := L_n(s) \otimes K_n(s)(D) \]

gives the probability that the pair \( (t, s') \) indicating the time of change and the new state will be a member of the Borel set \( D \in B(\mathbb{R}_+ \times S) \).

Now let \( (M^{(n)})_{n \in \mathbb{N}} \) be the projective system associated with \( L_n \otimes K_n : S \sim \mathbb{R}_+ \times S \) (formally, projective systems are associated with stochastic relations for which domain and range are identical; this is not the case for \( L_n \otimes K_n \), but we could resort to \( (L_n \otimes K_n)(t, s) := (L_n \otimes K_n)(s) \), thus we would be on the safe side. But things would look then quite awkward, so we close the eyes and let things as they are, hoping that things will work out — they will, just trust the author). Since \( M^{(n)} \) is defined inductively, we have in the inductive step for the Borel set \( D \subseteq (\mathbb{R}_+ \times S)^{n+1} \)

\[ M^{(n+1)}(s)(D) := \int_{(\mathbb{R}_+ \times S)^n} (L_{n+1}(s_n) \otimes K_{n+1}(s_n)) \left( D_{\langle t_0, s_1, \ldots, t_{n-1}, s_n \rangle} \right) \times \]

\[ \times M^{(n)}(s)(d(t_0, s_1, \ldots, t_{n-1}, s_n)). \]

We will need this formula later on.

Analyzing the expression, we see that at time \( n + 1 \) the probability that the pair of timing a transition and changing a state is an element of \( D_{\langle t_0, s_1, \ldots, t_{n-1}, s_n \rangle} \) is just

\[ (L_{n+1}(s_n) \otimes K_{n+1}(s_n)) \left( D_{\langle t_0, s_1, \ldots, t_{n-1}, s_n \rangle} \right), \]

provided the corresponding times and states that have been run through during times \( 1, \ldots, n \) is \( \langle t_0, s_1, \ldots, t_{n-1}, s_n \rangle \) which is captured through \( M^{(n)}(s)(d(t_0, s_1, \ldots, t_{n-1}, s_n)) \).

It is instructive to compute the integral with respect to the probability measure \( M^{(n)}(s) \) for a measurable and bounded map:

**Lemma 6.2.3** Let \( f : (\mathbb{R}_+ \times S)^n \rightarrow \mathbb{R} \) be measurable and bounded, then

\[ \int_{(\mathbb{R}_+ \times S)^n} f(t) \, dM^{(n)}(s) = \int_S \ldots \int_S \int_0^\infty \ldots \int_0^\infty f(t_0, s_1, \ldots, t_{n-1}, s_n) \times \]

\[ \times L_n(s_n) \, dt_{n-1} \, L_{n-1}(s_{n-2}) \, dt_{n-2} \ldots L_1(s) \, dt_0 \times \]

\[ \times K_n(s_{n-1}) \, ds_n \, K_{n-1}(s_{n-2}) \, ds_{n-1} \ldots K_1(s) \, ds_1. \]

**Proof** (Sketch) One first shows that the representation on the right hand side is true if \( f = \chi_A \) for some Borel set \( A \subseteq (\mathbb{R}_+ \times S)^n \). Correctness follows in this case from the definition, since

\[ \int_{(\mathbb{R}_+ \times S)^n} \chi_A \, dM^{(n)}(s) = M^{(n)}(s)(A). \]

The validity for indicator functions implies the validity for step functions, i.e., functions \( f \) of the form \( f = \sum_{i=0}^{n} \alpha_i \cdot \chi_{A_i} \) with Borel sets \( A_i \) and real \( \alpha_i \) through the integral’s additivity. One then observes that each non-negative bounded Borel map can be approximated by an increasing sequence of step functions from below, thus the equality is
true for non-negative \( f \) by the monotone convergence theorem (Proposition A.3.1). The general case writes \( f = f^+ + f^- \) with \( f^+(s) := \max\{f(s), 0\}, f^-(s) := \min\{f(s), 0\} \) and applies the previous case.

The last step rearranges the integrals according to the dependencies of their integration variables. This is admissible through Fubini’s Theorem on product integration, which permits interchanging the order of integration for product measures, in this case over the domains \( S \) and \( \mathbb{R}_+ \).

Now let \( M^\infty : S \rightsquigarrow (\mathbb{R}_+ \times S)^\infty \) be the projective limit of \( (M^{(n)})_{n \in \mathbb{N}} \), thus \( M^\infty(s)(A) \) is the probability for an infinite alternating path constructed from states and transition times that starts at state \( s \) to be in the Borel set \( A \in \mathcal{B}(\mathbb{R}_+ \times S)^\infty \). Sometimes a slightly different construction is used: it starts with an initial probability \( \pi \) on \( S \) and constructs a measure \( M^\pi_s \) on \( \mathcal{B}(S \times (\mathbb{R}_+ \times S)^\infty) \), only to specialize then \( \pi \) to \( \delta_s \). It is easy to see that \( M^\pi_s = M^\infty(s) \), and that vice versa

\[
M^\pi_s(A) = \int_S M^\infty(s)(A) \, \pi(ds)
\]

holds for \( A \in \mathcal{B}(\mathbb{R}_+ \times S)^\infty \). Which way to choose is a matter of taste and of convenience: The approach proposed here eases using tools from stochastic relations.

### Measurability on Paths

We will deal only with infinite paths. This is no loss of generality because events that happen at a certain time with probability 0 will have the effect that the corresponding infinite paths occur only with probability 0. Thus we do not prune the path; this makes the notation somewhat easier to handle without losing any substance.

The following Lemma looks innocent, but will turn out to be an important device:

**Lemma 6.2.4** \( \langle \sigma, t \rangle \mapsto \sigma \circ t \) is a Borel measurable map from \( \text{PATHS} \times \mathbb{R}_+ \) to \( S_\# \). In particular, the set \( \{ \langle \sigma, t \rangle \mid \sigma \circ t \in S \} \) is a measurable subset of \( \text{PATHS} \times \mathbb{R}_+ \).

Before we prove it, we need a simple auxiliary statement

**Lemma 6.2.5** Let \( (N, \mathcal{N}) \) be a measurable space, \( f : N \to \mathbb{R} \) be a Borel measurable map. Then

\[
\{ \langle n, x \rangle \mid f(n) > x \} \in \mathcal{N} \otimes \mathcal{B}(\mathbb{R}).
\]

**Proof** Put \( f_0(n, x) := \langle f(n), x \rangle \), then \( f_0 : N \times \mathbb{R} \to \mathbb{R} \) is \( \mathcal{N} \otimes \mathcal{B}(\mathbb{R}) \)-measurable. This is so since

\[
\mathcal{D} := \{ B \in \mathcal{B}(\mathbb{R} \times \mathbb{R}) \mid f_0^{-1}[B] \in \mathcal{N} \otimes \mathcal{B}(\mathbb{R}) \}
\]

is a \( \sigma \)-algebra, and since \( f_0^{-1}[B \times E] = f^{-1}[B] \times E \), hence we know that \( \mathcal{D} \) contains all measurable rectangles, thus \( \mathcal{D} = \mathcal{B}(\mathbb{R} \times \mathbb{R}) \).

Since \( \{ \langle n, x \rangle \mid f(n) > x \} = f_0^{-1}[L] \) with \( L := \{ \langle u, v \rangle \mid u > v \} \in \mathcal{B}(\mathbb{R} \times \mathbb{R}) \), the assertion is established. \( \dashv \)

The set \( \{ \langle n, s \rangle \mid f(n) > x \} \) may be visualized for \( N = \mathbb{R} \) as the area above the graph of \( f \).

**Proof of Lemma 6.2.4**

0. Note that we claim joint measurability in both components (which is strictly stronger than measurability in each component). Thus we have to show that \( \{ \langle \sigma, t \rangle \mid \sigma \circ t \in A \} \) is a measurable subset of \( \text{PATHS} \times \mathbb{R}_+ \), whenever \( A \subseteq S_\# \) is Borel.

---

194
1. Because for fixed \( i \in \mathbb{N} \) the map \( \sigma \mapsto \delta(\cdot, i) \) is a projection, \( \delta(\cdot, i) \) is measurable, hence \( \sigma \mapsto \sum_{i=0}^{j} \delta(\sigma, i) \) is. Consequently,

\[
\{ \langle \sigma, t \rangle \mid \sigma \cap t = \# \} = \{ \langle \sigma, t \rangle \mid \forall j : t \geq \sum_{i=0}^{j} \delta(\sigma, i) \}
\]

\[
= \bigcap_{j \geq 0} \{ \langle \sigma, t \rangle \mid t \geq \sum_{i=0}^{j} \delta(\sigma, i) \}.
\]

This is clearly a measurable set.

2. Put

\[
\text{Stop}(\sigma, t) := \inf \{ k \geq 0 \mid t < \sum_{i=0}^{k} \delta(\sigma, i) \},
\]

thus

\[
X_k := \{ \langle \sigma, t \rangle \mid \text{Stop}(\sigma, t) = k \} = \{ \langle \sigma, t \rangle \mid \sum_{i=0}^{k-1} \delta(\sigma, i) \leq t < \sum_{i=0}^{k} \delta(\sigma, i) \}
\]

is a measurable set by Lemma 6.2.5. Now let \( B \in \mathcal{B}(S) \) be a Borel set, then

\[
\{ \langle \sigma, t \rangle \mid \sigma \cap t \in B \} = \bigcup_{k \geq 0} \{ \langle \sigma, t \rangle \mid \sigma \cap t \in B, \text{Stop}(\sigma, t) = k \}
\]

\[
= \bigcup_{k \geq 0} \{ \langle \sigma, t \rangle \mid \sigma[k] \in B, \text{Stop}(\sigma, t) = k \}
\]

\[
= \bigcup_{k \in \mathbb{N}} \left( X_k \cap \left( \prod_{i<k} (S \times \mathbb{R}_+) \times (B \times \mathbb{R}_+) \times \prod_{i>k} (S \times \mathbb{R}_+) \right) \right).
\]

Because \( X_k \) is measurable, the latter set is measurable. This establishes measurability of the \( \cap \)-map. \( \dashv \)

As a consequence, we obtain the measurability of some sets and maps which will be important for the later development. But let us just agree on a notational convention for improving readability: the letter \( \sigma \) will always denote a generic element of PATHS, and the letter \( \tau \) always a generic element of \( \mathbb{R}_+ \times \text{PATHS} \).

**Proposition 6.2.6** We observe the following properties:

1. \( \{ \sigma \mid \sum_{i=0}^{\infty} \delta(\sigma, i) \text{ exists and is finite} \} \) is a measurable subset of PATHS,

2. \( \{ \langle \sigma, t \rangle \mid \lim_{i \to \infty} \delta(\sigma, i) = t \} \) is a measurable subset of \( \text{PATHS} \times \mathbb{R}_+ \),

3. both

\[
s \mapsto \lim_{t \to \infty} M^\infty(s)(\{ \tau \mid \langle s, \tau \rangle \cap t \in A \})
\]

and

\[
s \mapsto \lim_{t \to \infty} \sup M^\infty(s)(\{ \tau \mid \langle s, \tau \rangle \cap t \in A \})
\]

are measurable maps \( X \to \mathbb{R}_+ \) for each Borel set \( A \subseteq S \).
\textbf{Proof} 0. The proof makes crucial use of the fact that the real line is a complete metric space (so each Cauchy sequence converges), and that the rational numbers are dense, forming a countable set.

1. Since $\sum_{i=0}^{n} \delta(\sigma, i)$ exists and is finite iff given $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that $|\sum_{i=0}^{n} \delta(\sigma, i)| < \epsilon$ whenever $n_1, n_2 \geq n$, we see that

$$\{\sigma \mid \sum_{i=0}^{n} \delta(\sigma, i) \text{ exists and is finite} \} = \bigcap_{Q_1 \epsilon > 0} \bigcup_{n \in \mathbb{N}} \bigcup_{n_1, n_2 \geq n} \{\sigma \mid |\sum_{i=0}^{n} \delta(\sigma, i)| < \epsilon\}.$$ 

Measurability of $\sigma \mapsto \delta(\sigma, i)$ for each $i$ follows from Lemma 6.2.4. This implies measurability of the set in part 1, since the union and the intersections are defined over countable index sets.

2. The same argument applies basically to set of all paths and times the timing labels converge to in part 2:

$$\{\langle \sigma, t \rangle \mid \lim_{i \to \infty} \delta(\sigma, i) = t\} = \bigcap_{Q_2 \epsilon > 0} \bigcup_{n \in \mathbb{N}} \bigcup_{m \geq n} \{\langle \sigma, t \rangle \mid |\delta(\sigma, m) - t| < \epsilon\}.$$ 

By Lemma 6.2.4, the set

$$\{\langle \sigma, t \rangle \mid |\delta(\sigma, m) - t| < \epsilon\} = \{\langle \sigma, t \rangle \mid \delta(\sigma, m) > t - \epsilon \} \cap \{\langle \sigma, t \rangle \mid \delta(\sigma, m) < t + \epsilon\}$$

is a measurable subset of $\text{PATHS} \times \mathbb{R}_+$, and since the union and the intersections are countable, measurability is inferred.

3. From the definition of the $@$-operator it is immediate that given an infinite path $\sigma$ and a time $t \in \mathbb{R}_+$, there exists a rational $t'$ with $\sigma@t = \sigma@t'$. Thus we obtain for an arbitrary real number $x$, an arbitrary Borel set $A \subseteq S$ and $s \in S$

$$\liminf_{t \to \infty} M^\infty(s)(\{\tau \mid \langle s, \tau \rangle@t \in A\}) \leq x \iff \sup_{t \geq 0} \inf_{r \geq t} M^\infty(s)(\{\tau \mid \langle s, \tau \rangle@r \in A\}) \leq x$$

$$\iff s \in \bigcap_{Q \geq 0} \bigcup_{Q \geq t} A_{r,x}$$

with

$$A_{r,x} := \{s' \mid M^\infty(s')(\{\tau \mid \langle s', \tau \rangle@r \in A\}) \leq x\}.$$ 

We infer that $A_{r,x}$ is a measurable subset of $S$ from the fact that $M^\infty$ is a stochastic relation and from Lemma 6.2.4, cp. Lemma A.3.5. Since a map $f : W \to \mathbb{R}$ is measurable iff each of the sets $\{w \in W \mid f(w) \leq s\}$ is a measurable subset of $W$, the assertion follows for the first map in part 1. The second part is established in exactly the same way, using that $f : W \to \mathbb{R}$ is measurable iff $\{w \in W \mid f(w) \geq s\}$ is a measurable subset of $W$, and observing

$$\limsup_{t \to \infty} M^\infty(s)(\{\tau \mid \langle x, \tau \rangle@t \in A\}) \geq x \iff \inf_{Q \geq 0} \sup_{Q \geq t} M^\infty(s)(\{\tau \mid \langle s, \tau \rangle@r \in A\}) \geq x.$$ 

As a consequence we obtain that the set on which the asymptotic behavior of the transition times is reasonable in the sense that it tends probabilistically to a limit is well behaved in terms of measurability:

196
Corollary 6.2.7 Let $A \subseteq X$ be a Borel set, then

1. the set $Q_A := \{s \in S \mid \lim_{t \to \infty} M^\infty(s)(\{\tau \mid (s, \tau) \in A\}) \text{ exists}\}$ on which the limit exists is a Borel subset of $S$,

2. $s \mapsto \lim_{t \to \infty} M^\infty(s)(\{\tau \mid (s, \tau) \in A\})$ is a measurable map $Q_A \to \mathbb{R}_+$.

Proof Since $s \in Q_A$ iff

$$\liminf_{t \to \infty} M^\infty(x)(\{\tau \mid (s, \tau) \in A\}) = \limsup_{t \to \infty} M^\infty(x)(\{\tau \mid (s, \tau) \in A\}),$$

and since the set on which two Borel measurable maps coincide is a Borel set itself, the first assertion follows from Proposition 6.2.6, part 3. This implies the second assertion. $\dashv$

We have seen that the set

$$Z := \{\sigma \mid \sum_{i \geq 0} \delta(\sigma, i) \text{ exists and is finite}\}$$

of all Zeno paths is measurable in the universe of our paths. Math 101 tells us that $Z \subseteq C$ with $C$ as the set of all paths the transition times of which tend to zero, thus

$$C := \{\sigma \mid \lim_{i \to \infty} \delta(\sigma, i) = 0\}.$$ 

By Proposition 6.2.6 this is also a measurable set. We will show now that a Zeno path will only occur with probability 0, provided the probabilities $L_n$ that govern the transitions do not concentrate their mass close to zero (this means that very short transition times will occur quite infrequently).

We establish this property of Zeno paths under the assumption that the relations $(L_n)_{n \in \mathbb{N}}$ are uniformly bounded at the origin, but we treat the problem a wee bit more general.

Definition 6.2.8 The sequence $L_n : S \to \mathbb{R}_+$ is called uniformly bounded at time $t$ iff given $\epsilon > 0$ there is $\delta > 0$ such that

$$\sup_{s \in S} L_n(s)([t - \delta, t + \delta] \cap \mathbb{R}_+) < \epsilon$$

for all but a finite number of indices $n$.

Thus uniformly bounded at $t$ means for $(L_n)_{n \in \mathbb{N}}$ that no positive mass is associated with $t$, uniformly for all states and almost all $n$. Thus we cannot say that $t$ is a time in which state changes occur with positive probability for any state and for almost all $n$. By the way, the condition can be formulated in terms of limits as

$$\forall \epsilon > 0 \exists \delta > 0 : \limsup_{n \to \infty} \sup_{s \in S} L_n(s)([t - \delta, t + \delta] \cap \mathbb{R}_+) < \epsilon.$$ 

Observation 6.2.9 Assume that we work in the rate model. Let as in Observation 6.2.2

$$L_n(s)(F) := \frac{1}{E(s)} \cdot \int_F e^{-E(s) \cdot t} dt,$$
and assume that $\rho := \sup_{s \in S} R(s)(S)$ is finite. Then the sequence $(L_n)_{n \in \mathbb{N}}$ is uniformly bounded at each time $t$. This is so since

$$L_n(s)([t_1, t_2]) = e^{-E(s)\cdot t_1} \cdot (1 - e^{-E(s)\cdot (t_2 - t_1)}) \leq 1 - e^{-\rho (t_2 - t_1)}.$$ 

This difference, which is independent of state $x$, can be brought arbitrarily close to 0. ▶

Thus we will prove a slightly more general statement that in [22, Theorem 3.2]:

**Proposition 6.2.10** The following holds for each $s \in S$

1. $M^\infty(s)(\{ \tau \mid \lim_{i \to \infty} \delta(\sigma, i) = t \}) = 0$, provided $(L_n)_{n \in \mathbb{N}}$ is uniformly bounded at $t$,

2. $M^\infty(s)(\{ \tau \mid \sum_{i \geq 0} \delta(\langle s, \tau \rangle, i) \text{ exists and is finite} \}) = 0$, provided $(L_n)_{n \in \mathbb{N}}$ is uniformly bounded at the origin. Consequently, the set of all Zeno paths is negligible.

**Proof** 0. The remarks preceding Definition 6.2.8 imply that we only have to establish part 1, part 2 will then follow immediately for $t = 0$. Fix $t \in \mathbb{R}_+$ and assume that $(L_n)_{n \in \mathbb{N}}$ is uniformly bounded at $t$.

1. Since

$$\lim_{i \to \infty} \delta(\sigma, i) = t \iff \forall \epsilon > 0 \exists n \in \mathbb{N} \forall k \geq n : |\delta(\sigma, j) - t| < \epsilon,$$

we can represent $C$ as

$$C = \bigcap_{Q : \exists \epsilon > 0} \bigcup_{n \in \mathbb{N}} N_{\epsilon, n}$$

with

$$N_{r, n} := \{ \sigma \mid |\delta(\sigma, n + j) - t| < r \text{ for all } j \in \mathbb{N} \}.$$

It is clear from the definition that

$$N_{r, n} = \bigcap_{k \geq 0} N'_{r, n, k},$$

where

$$N'_{r, n, k} := \{ \sigma \mid |\delta(\sigma, n + j) - t| < r \text{ for } 0 \leq j \leq k \} = N''_{r, n, k} \times \prod_{j > n + k} (\mathbb{R}_+ \times X).$$

The sequence $(N'_{r, n, k})_{k \in \mathbb{N}}$ is monotonically decreasing, hence

$$M^\infty(s)(\{ \tau \mid \langle s, \tau \rangle \in N_{r, n} \}) = \inf_{k \in \mathbb{N}} M^\infty(s)(\{ \tau \mid \langle s, \tau \rangle \in N'_{r, n, k} \}).$$

From the construction of $M^\infty(s)$ as projective limit of $(M^{(n)})_{n \in \mathbb{N}}$ we see that

$$M^\infty(s)(\{ \tau \mid \langle s, \tau \rangle \in N'_{r, n, k} \}) = M^{(n+k)}(s)(\{ \tau \mid \langle \langle s, \tau \rangle \rangle, \tau \rangle \in N''_{r, n, k} \}),$$
which by Lemma 6.2.3 may be evaluated as

\[
M^{(n+k)}(s)\{\{\tau \mid \langle s, \tau \rangle \in N''_{n,n,k}\}\} = \\
\int S \ldots \int S L_{n+k}(s_{n+k-1})([t-r,t+r] \cap \mathbb{R}+) \cdots \cdot L_n(s_{n-1})([t-r,t+r] \cap \mathbb{R}+) \times \\
\times K_{n+k-1}(s_{n+k-2})(ds_{n+k-1}) \cdots K_2(s_1)(dx_2)K_1(s)(ds_1).
\]

Well, that’s not too bad.
2. Now if \(0 < \epsilon < 1\) is given, so we can find \(\eta > 0\) such that

\[ L_n(s)([t-\eta, t+\eta] \cap \mathbb{R}+) < \epsilon \]

is true for all \(n \in \mathbb{N}\) and all \(s \in S\) due to \((L_n)_{n \in \mathbb{N}}\) being uniformly bounded at the origin. Consequently, \(M^{(n+k)}(s)\{\{\tau \mid \langle s, \tau \rangle \in N''_{n,n,k}\}\} \leq \epsilon^{k+1}\), which implies

\[ M^\infty(s)\{\{\tau \mid \langle x, \tau \rangle \in N_{\eta,n}\}\} \leq \inf_{k \in \mathbb{N}} \epsilon^k = 0. \]

But this trivially implies \(M^\infty(s)\{\{\tau \mid \lim_{i \to \infty} \delta(\sigma, i) = t\}\} = 0\). 

If we assume that \((L_n)_{n \in \mathbb{N}}\) is uniformly bounded at each point in time, then the mass associated with each \(L_n(s)\) is not concentrated at any time \(t\), for otherwise the probability of hitting arbitrary small intervals \([t-\delta, t+\delta]\) around \(t\) cannot be made arbitrary small. Thus there is no preferred, pointed timing behavior.

**Interpreting The Logic**

Now that we know how to probabilistically describe the behavior of paths, we are ready for a probabilistic interpretation of CSL. This is done using the sequences \((K_n)_{n \in \mathbb{N}}\) and \((L_n)_{n \in \mathbb{N}}\), from which the stochastic relation \(M : X \sim \mathbb{R}_+ \times \text{PATHS}\) has been constructed. The interpretations for the formulas are established, and we show that the sets of states resp. paths on which formulas are valid are Borel measurable.

To get started on the formal definition of the semantics, we assume that we know for each atomic proposition which state it is satisfied in, so we fix a map \(L\) that maps \(P\) to \(B(S)\), assign each atomic proposition a Borel set of states. The semantics is described as usual recursively through relation \(\models\) between states resp. paths, and formulas as follows:

1. \(s \models \top\) is true for all \(s \in S\).
2. \(s \models a\) iff \(s \in L(a)\).
3. \(s \models \varphi_1 \land \varphi_2\) iff \(s \models \varphi_1\) and \(s \models \varphi_2\).
4. \(s \models \neg \varphi\) iff \(s \models \varphi\) is false.
5. \(s \models S_{\text{loc}}(\varphi)\) iff \(\lim_{t \to \infty} M^\infty(s)\{\{\tau \mid \langle s, \tau \rangle \models t \models \varphi\}\}\) exists and is \(\triangleright p\).
6. \(s \models P_{\text{loc}}(\psi)\) iff \(M^\infty(s)\{\{\tau \mid \langle x, \tau \rangle \models \psi\}\}\ \triangleright p\).
7. \(\sigma \models \mathcal{X}^l \varphi\) iff \(\sigma[1] \models \varphi\) and \(\delta(\sigma, 0) \in I\).
8. \( \sigma \models \varphi_1 U^I \varphi_2 \) iff \( \exists t \in I : \sigma@t \models \varphi_2 \) and \( \forall t' \in [0,t] : \sigma@t' \models \varphi_1 \).

Denote by \([\varphi]\) again the set of all states for which the state formula \( \varphi \) holds, resp. the set of all paths for which the path formula \( \varphi \) is valid. We do not distinguish notationally between these sets, as far as the basic domains are concerned, since it should always be clear whether we describe a state formula or a path formula.

We show that we are dealing with measurable sets. The until-operator requires some attention, thus we single it out, before diving into a general discussion on measurability again.

**Lemma 6.2.11** Assume that \( A_1 \) and \( A_2 \) are Borel subsets of \( S \), and \( I \subseteq \mathbb{R}_+ \) be an interval, then

\[
U(I, A_1, A_2) := \{ \sigma \mid \exists t \in I : \sigma@t \in A_2 \land \forall t' \in [0,t] : \sigma@t' \in A_1 \}
\]

is a measurable set of paths, thus \( U(I, A_1, A_2) \in \mathcal{B}(\text{PATHS}) \).

**Proof 0.** Remember that, given a path \( \sigma \) and a time \( t \in \mathbb{R}_+ \) there exists a rational time \( t_r \leq t \) with \( \sigma@t = \sigma@t_r \). Consequently,

\[
U(I, A_1, A_2) = \bigcup_{t \in \mathbb{Q} \cap I} \left( \{ \sigma \mid \sigma@t \in A_1 \} \cap \bigcap_{t' \in \mathbb{Q} \cap [0,t]} \{ \sigma \mid \sigma@t' \in A_2 \} \right).
\]

The inner intersection is countable and is performed over measurable sets by Lemma 6.2.4, thus forms a measurable set of paths. Intersecting it with a measurable set and forming a countable union yields a measurable set again. \( \dashv \)

This is the crucial step towards establishing

**Proposition 6.2.12** The set \([\varphi]\) is Borel, whenever \( \varphi \) is a state formula or a path formula.

**Proof 0.** The proof proceeds by induction on the structure of the formula \( \varphi \). The induction starts with the formula \( \top \), for which the assertion is true, and with the atomic propositions, for which the assertion follows from the assumption on \( L \): \([a] = L(a) \in \mathcal{B}(X) \). We assume for the induction step that we have established that \([\varphi],[\varphi_1]\) and \([\varphi_2]\) are Borel measurable.

1. For the next-operator we write

\[
[\begin{array}{c}
\chi^I \varphi \\
\end{array}] = \{ \sigma \mid \sigma[1] \in [\varphi] \text{ and } \delta(\sigma,0) \in I \}.
\]

This is the cylinder set \((S \times I \times [\varphi] \times \mathbb{R}_+) \times \text{PATHS}, \) hence is a Borel set.

2. The until-operator may be represented through

\[
[\varphi_1 U^I \varphi_2] = U(I, [\varphi_1],[\varphi_2]),
\]

which is a Borel set by Lemma 6.2.11.

3. Since \( M^\infty : S \rightsquigarrow (\mathbb{R}_+ \times S)^\infty \) is a stochastic relation, we know that

\[
[P_{\varphi_p}(\psi)] = \{ s \in S \mid M^\infty(s)(\{ \tau \mid (s, \tau) \in [\varphi] \}) \models p \}
\]

is a Borel set.
6.3 Bisimulations for CSL

4. We know from Corollary 6.2.7 that the set

\[ Q_\varphi := \{ s \in S \mid \lim_{t \to \infty} M^\infty(s)(\{ \tau \mid (s, \tau)@t \in [\varphi] \}) \text{ exists} \} \]

is a Borel set, and that

\[ \ell_\varphi : Q_\varphi \ni s \mapsto \lim_{t \to \infty} M^\infty(s)(x) (\{ \tau \mid (s, \tau)@t \in [\varphi] \}) \in [0, 1] \]

is a Borel measurable function. Consequently,

\[ [S_{\geq 0.5}(\varphi)] = \{ s \in Q_\varphi \mid \ell_\varphi(s) \geq 0.5 \} \]

is a Borel set. ⊣

Measurability of the sets on which a given formula is valid is of course a prerequisite for computing interesting properties. So we can compute e.g.

\[ P_{\geq 0.5}((-\text{down}) U^{10,20}[S_{\geq 0.8}(\text{up}_2 \lor \text{up}_3))] \]

as the set of all states that with probability at least 0.5 will reach a state between 10 and 20 time units so that the system is operational (\( \text{up}_2, \text{up}_3 \in P \)) in a steady state with a probability of at least 0.8; prior to reaching this state, the system must be operational continuously (\( \text{down} \in P \)).

6.3 Bisimulations for CSL

We will assume for the rest of this section that both \( K_n \) and \( L_n \) are independent of \( n \), so that we work with \( K \) resp. \( L \) instead; the projective limit will be denoted by \( M \) rather than by \( M^\infty \), reminding the reader of this assumption. Thus the probabilities for a transition and those governing the time for staying in a state are independent of the step in which we are considering the system. This assumption includes the usual assumptions in which the transition probabilities and the probabilities for timing are derived from a rate function independent of the step in which the game is performed.

As a first consequence of making the basic probabilities independent of step \( n \) we obtain a recursive formulation for the transition law \( M : X \rightsquigarrow (\mathbb{R}_+ \times S)\infty \) that reflects the domain equation \( (\mathbb{R}_+ \times S)\infty = (\mathbb{R}_+ \times S) \times (\mathbb{R}_+ \times X)\infty \).

**Lemma 6.3.1** If \( D \in \mathcal{B}((\mathbb{R}_+ \times S)\infty) \), then

\[ M(s)(D) = \int_{\mathbb{R}_+ \times S} M(s')(D_{\langle t, s' \rangle}) M^{(1)}(s)(d\langle t, s' \rangle) \]

holds for all \( s \in S \).

**Proof** Recall that \( D_{\langle t, s' \rangle} = \{ \tau \mid \langle t, s', \tau \rangle \in D \} \). Let

\[ D = (H_1 \times \ldots \times H_{n+1}) \times \prod_{j > n}(\mathbb{R}_+ \times X) \]
be a cylinder set with \( H_i \in B(\mathbb{R}_+ \times S) \), \( 1 \leq i \leq n + 1 \). The equation in question boils in this case down to

\[
M^{(n+1)}(s)(H_1 \times \ldots \times H_{n+1}) = \int_{H_1} M^{(n)}(s')(H_2 \times \ldots \times H_{n+1})M^{(1)}(s)(d(t, x')).
\]

This may easily be derived from Lemma 6.2.3. Consequently, the equation in question holds for all cylinder sets, thus the \( \pi\lambda\)-Theorem (Proposition A.1.1) implies that it holds for all Borel subsets of \((\mathbb{R}_+ \times S)^\infty\).

This decomposition indicates that we may first select a state \( s \) a new state and a transition time; with these data the system then works just as if the selected new state would have been the initial state. New states and transition times are being averaged over, since we select these items according to a probability law. Lemma 6.3.1 may accordingly be interpreted as a Markov property for a process the behavior of which is independent of the specific step that is undertaken.

### 6.3.1 Definition and Properties of \( \rho_F \)

Returning to the logic, fix a set \( F \) of state formulas, and define the central equivalence relation

\[
s \rho_F s' \iff \forall \varphi \in F : [s \models \varphi \iff s' \models \varphi],
\]

then \( \rho_F \) is smooth due to \( F \) being countable. We will investigate in this section the equivalence \( \rho_F \). First, the closure \( \text{wrap}(F) \) of \( F \) will be defined as the smallest set of formulas containing \( F \) and being closed under the logic’s operators, and it will be investigated under which conditions \( \rho_{\text{wrap}(F)} = \rho_F \) holds. An answer to this question makes life easier, since testing satisfaction only on \( F \) is presumably much easier than testing on \( \text{wrap}(F) \), in particular when \( F = P \) (so that \( \text{wrap}(F) = \mathcal{L}_P \)). We will examine an enabling condition, using smooth equivalence relations and congruences as the decisive tool. This leads to a discussion of bisimulations, the results obtained for congruences will be transported for an investigation of bisimilar states. Conditions under which \( P \)-bisimilarity and the satisfaction of the same formulas will be formulated at the end of this section.

The closure \( \text{wrap}(F) \) of \( F \) is defined as the smallest set of formulas in \( \mathcal{L}_P \) which contains \( F \) and which is closed under the defining operations for the logic. Formally, \( \text{wrap}(F) \) is the set of all \( F \)-state formulas which are defined through the following rules:

- **F-state formulas** are defined through the syntax

\[
\varphi ::= \top | \Phi | \neg \varphi | \varphi \land \varphi' | S_{=p}(\varphi) | P_{=p}(\psi).
\]

Here \( \Phi \in F \) is a formula in \( F \), \( \psi \) is an \( F \)-path formula, \( \triangleleft \) is one of the relational operators \( <, \leq, \geq, > \), and \( p \in [0, 1] \) is a rational number.

- **F-path formulas** are defined through

\[
\psi ::= X^I \varphi \mid \varphi U^I \varphi'
\]

with \( \varphi, \varphi' \) as \( F \)-state formulas, \( I \subseteq \mathbb{R}_+ \) a closed interval of the real numbers with rational bounds.
Thus we start in building up \( F \)-formulas from elements of \( F \) as the base, just as we started building up \( \mathcal{L}_P \) from the set \( P \) of atomic propositions. Observe that \( \text{wrap}(P) = \mathcal{L}_P \). We will investigate the smooth relations \( \rho_F \) and \( \rho_{\text{wrap}}(F) \) and will establish that under a mildly restrictive condition \( \rho_F = \rho_{\text{wrap}}(F) \) holds. This result looks rather modest, but it has some interesting consequences in terms of bisimulations. They will be discussed after the proof.

The mild condition that will enable us to establish the relations’ equality was detected by Desharnais and Panagaden in [22] for the fragment of CSL investigated there.

**Definition 6.3.2** A set \( F \) of formulas is said to satisfy the DP-condition iff \( F \) has these properties: \( F \) is closed under conjunctions, and \( \mathcal{P}_{\leq}(\mathcal{X}^{I} \varphi) \in F \) whenever \( \varphi \in F, p \in [0, 1] \) is rational, and \( I \subseteq \mathbb{R}_+ \) is a closed interval with rational endpoints.

We will see that the closedness under conjunction will later enable us to apply the \( \pi\lambda \)-Theorem for making sure that a condition carries over from the set of generators (in this case \( \{[[\varphi]] \mid \varphi \in F\} \)) to the \( \sigma \)-algebra generated from it. Closedness under the next operator will have a special consequence, as we will see in a moment.

The DP-condition makes sure that the probabilities for a transition of \( \rho_F \)-equivalent states into a state in which a formula in \( F \) is valid are identical. This is quite comparable to the observation one makes for stochastic Kripke models for modal logics: there it is well known that the probabilities for making a move into a state in which the same formula is satisfied after an action coincide for equivalent states as well, see Lemma 6.1.18.

**Lemma 6.3.3** If \( s \), \( \rho_F s' \) and \( \varphi \in F \), then \( K(s)([[\varphi]]) = K(s')([[\varphi']]) \), provided \( F \) satisfies the DP-condition.

**Proof** Suppose that we find for \( s \), \( \rho_F s' \) a formula \( \varphi' \in F \) such that

\[
K_1(s)([[\varphi']]) < r \leq K_1(s')([[\varphi']]),
\]

where \( r \) may be assumed to be rational. Since

\[
\{ \tau \mid \langle s, \tau \rangle \models \mathcal{X}^{\mathbb{R}_+} \varphi' \} = (\mathbb{R}_+ \times [[\varphi']]) \times (\mathbb{R}_+ \times S)^\infty,
\]

we conclude that

\[
K_1(s)([[\varphi']]) = M(s)(\{ \tau \mid \langle s, \tau \rangle \models \mathcal{X}^{\mathbb{R}_+} \varphi' \}).
\]

But this implies that \( s \models \mathcal{P}_{\leq}(\mathcal{X}^{\mathbb{R}_+} \varphi') \), similarly, \( s' \not\models \mathcal{P}_{\leq}(\mathcal{X}^{\mathbb{R}_+} \varphi') \). But the DP-condition implies that \( \mathcal{P}_{<}(\mathcal{X}^{\mathbb{R}_+} \varphi') \in F \), which is a contradiction. \( \dashv \)

This Lemma is actually a first step towards establishing that \( \rho_F \) generates a congruence for \( M \). It requires an extension of the equivalence relation \( \rho_F \) on \( S \) to \( (\mathbb{R}_+ \times S)^\infty \). The basic idea is to relate the alternating states in such a sequence through \( \rho_F \), and to leave the residence times alone, which means to relate them through the identity relation \( \Delta_{\mathbb{R}_+} \).

Thus \( \langle t_0, s_1, t_1, \ldots \rangle \) will be related to \( \langle t'_0, s'_1, t'_1, \ldots \rangle \) iff \( s_i \rho_F s'_i \) and \( t_i = t'_i \) for all indices \( i \).

In view of Lemma 5.1.18 we form the product relation \( \times_{n \in \mathbb{N}} (\Delta_{\mathbb{R}_+} \times \rho_F) = (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty \).

To alleviate the heavy notation somewhat, we abbreviate

\[
\rho_F^{(n)} := (\Delta_{\mathbb{R}_+} \times \rho_F)^n, \\
\rho_F^{(\infty)} := (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty.
\]
Proposition 6.3.4 Assume that $F$ satisfies the DP-condition, then

$$c_F := (\rho_F, \rho_F^{(\infty)})$$

is a congruence for $M : S \rightsquigarrow (\mathbb{R}_+ \times S)^\infty$.

Proof 0. We need to show that

$$(\dagger) M(s)(D) = M(s')(D),$$

holds for each $\rho_F^{(\infty)}$-invariant Borel set $D$, provided $s \rho_F s'$ holds. We know from Lemma 5.1.18 that

$$INV \left( B((\mathbb{R}_+ \times S)^\infty), \rho_F^{(\infty)} \right) = \bigotimes_{n \in \mathbb{N}} INV \left( B(\mathbb{R}_+ \times S), \Delta_{\mathbb{R}_+} \times \rho_F \right)$$

holds, and from the construction of the infinite product of measurable spaces we see that we may restrict our attention to cylinder sets the factors of which are $\Delta_{\mathbb{R}_+} \times \rho_F$-invariant. But since $INV (B(\mathbb{R}_+ \times S), \Delta_{\mathbb{R}_+} \times \rho_F)$ is generated by $\{I \times [\varphi] \mid I \subseteq \mathbb{R}_+ \text{ is an interval}, \varphi \in F\}$, it is sufficient for establishing Eq. (\dagger) that the equation

$$(\dagger') M^{(n)}(s) ((I_1 \times [\varphi_1]) \times \ldots \times (I_n \times [\varphi_n])) = M^{(n)}(s') ((I_1 \times [\varphi_1]) \times \ldots \times (I_n \times [\varphi_n]))$$

holds, whenever $s \rho_F s'$, where $I_1, \ldots, I_n$ are intervals in $\mathbb{R}_+$ with rational endpoints and $\varphi_1, \ldots, \varphi_n$ are formulas in $F$. This is done by induction on $n$.

Fix $s, s'$ with $s \rho_F s'$, intervals $(I_n)_{n \in \mathbb{N}}$ in $\mathbb{R}_+$ with rational endpoints, and formulas $(\varphi_n)_{n \in \mathbb{N}}$ in $F$, and put $B_n := [\varphi_n]$ as the set of states in which $\varphi_n$ is valid.

1. The induction starts at $n = 1$ with the observation that

$$M^{(1)}(s)(I_1 \times B_n) = M(s)(\{\tau \mid \langle s, \tau \rangle \in I_1 \times B_1\} = M(s)(\{\tau \mid \langle s, \tau \rangle \models \mathcal{X}^{I_1} \varphi_1\}.$$

Thus we have for an arbitrary rational $p$

$$M^{(1)}(s)(I_1 \times B_1) \leq p \iff s \models \mathcal{P}_{\omega_0} (\mathcal{X}^{I_1} \varphi_1) \iff s' \models \mathcal{P}_{\omega_0} (\mathcal{X}^{I_1} \varphi_1) \quad (\text{since } s \rho_F s') \iff M^{(1)}(s')(I_1 \times B_1) \leq p.$$

Consequently, $M^{(1)}(s)(I_1 \times B_1) = M^{(1)}(s')(I_1 \times B_1)$ is established.

2. Assume for the induction step that the assertion is true for $n$. This implies in particular that $(\rho_F, \rho_F^{(n)})$ is a congruence for $M^{(n)} : S \rightsquigarrow (\mathbb{R}_+ \times S)^n$. From the Markov property in Lemma 6.3.1 we infer that

$$M^{(n+1)}(s)((I_1 \times B_1) \times \cdots \times (I_{n+1} \times B_{n+1})) =$$

$$\int_{I_1 \times B_1} M^{(n)}(y)((I_2 \times B_2) \times \cdots \times (I_{n+1} \times B_{n+1})) M^{(1)}(s)(d(t, y)) =$$

$$\int_{\mathbb{R}_+ \times S} \chi_{I_1 \times B_1}(t, y) \cdot M_n(y)((I_2 \times B_2) \times \cdots \times (I_{n+1} \times B_{n+1})) M^{(1)}(s)(d(t, y))$$

204
The intermediate goal is to prove that induction step through the integral representation rendering the Markov property. Reflecting upon the proof, we see that the DP-condition on $F$ which implies equation (\ref{eq:dp-condition}). We observe the following properties:

We define auxiliary operators: let $A, A_1, A_2$ be subsets of $X$, $B$ be a subset of $\text{PATHS}$, and $I \subseteq \mathbb{R}_+$ an interval with rational bounds, then

\[
P_{\text{op}}(B) := \{s \in S \mid M(s)(\{\tau \mid \langle s, \tau \rangle \in B\}) \propto p\}
\]
\[
Q_A := \{s \in S \mid \lim_{t \to \infty} M(s)(\{\tau \mid \langle s, \tau @ t \in A\}) \text{ exists}\}
\]
\[
f_A(s) := \lim_{t \to \infty} M(s)(\{\tau \mid \langle s, \tau \rangle \in A\}), \text{ if } s \in Q_A
\]
\[
S_{\text{op}}(A) := \{s \in Q_A \mid f_A(s) \propto p\}
\]
\[
X(I, A) := \{\sigma \mid \sigma[1] \in A \land \delta(\sigma, 0) \in I\}.
\]

We observe the following properties:

\textbf{Lemma 6.3.5} Let $F$ be a set of formulas, and recall that $\rho_F \times \Delta_{\mathbb{R}_+}$ denotes the smooth equivalence relation

\[
\langle s, t \rangle (\rho_F \times \Delta_{\mathbb{R}_+}) \langle s', t' \rangle \iff s \rho_F s' \land t = t',
\]

on $S \times \mathbb{R}_+$. Assume that $F$ satisfies the DP-condition. We observe the following properties:

1. If $B \in \mathcal{INV}(\mathcal{B}(\text{PATHS}), (\rho_F \times \Delta_{\mathbb{R}_+})^\infty)$, then $P_{\text{op}}(B) \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$.

2. If $A \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$, then $Q_A \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$, $S_{\text{op}}(A) \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$, and $X(I, A) \in \mathcal{INV}(\mathcal{B}(\text{PATHS}), (\rho_F \times \Delta_{\mathbb{R}_+})^\infty)$.

3. If $A_1, A_2 \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$, then $U(I, A_1, A_2) \in \mathcal{INV}(\mathcal{B}(\text{PATHS}), (\rho_F \times \Delta_{\mathbb{R}_+})^\infty)$.
Proof 1. Since $F$ satisfies the CP-condition, we know from Proposition 6.3.4 that $c_F$ is a congruence for $M : S \rightsquigarrow (\mathbb{R}_+ \times S)\infty$. From Lemma 5.2.5 and Corollary 2.2.10 we infer that $s \mapsto M(s)(B_s)$ is a $\mathcal{INV}(B(S), \rho_F)\cdot B(\mathbb{R}_+)$-measurable function, where $B_s = \{ \tau \mid \langle s, \tau \rangle \in B \}$. This implies the assertion in part 1.

2. Define for $t \in \mathbb{R}_+$ the set $J_A := \{ \sigma \mid \sigma @ t \in A \}$, then $J_A \in \mathcal{INV}(B(\text{PATHS}), (\rho_F \times \Delta_{\mathbb{R}_+})\infty)$. In fact, suppose $\sigma (\rho_F \times \Delta_{\mathbb{R}_+})\infty \sigma'$, then $\delta(\sigma, i) = \delta(\sigma', i)$ holds for all $i$ (this is so since the equivalence does not affect the timing information), thus $\text{Stop}(\sigma, r) = \text{Stop}(\sigma', r)$ for all $r \geq 0$. Consequently, we obtain (cf. the proof for Lemma 6.2.4)

\[
\sigma \in J_A \iff \sigma @ t \in A
\]

\[
\iff \exists k : \text{Stop}(\sigma, t) = k, \sigma[k] \in A
\]

\[
\iff \exists k : \text{Stop}(\sigma', t) = k, \sigma[k] \in A
\]

\[
\iff \sigma' \in J_A,
\]

establishing the invariance of $J_A$. Clearly, $J_A$ is a Borel set by Lemma 6.2.4.

Again, we infer that $s \mapsto M(s)(\{ \tau \mid \langle s, \tau \rangle \in J_A \})$ is $\mathcal{INV}(B(S), \rho_F)\cdot B(\mathbb{R}_+)$-measurable, hence

\[
A_{t,s} := \{ s' \mid M(s)(\{ \tau \mid \langle s', \tau \rangle @ t \in A \}) \leq s \}
\]

defines an $\rho_F$-invariant Borel set. We know from the proof of Proposition 6.2.6 that

\[
\liminf_{t \to \infty} M(s)(\{ \tau \mid \langle s, \tau \rangle @ t \in A \}) \leq x \iff s \in \bigcap_{Q \ni t} \bigcup_{Q \ni t} A_{t,s},
\]

thus

\[
s \mapsto \liminf_{t \to \infty} M(s)(\{ \tau \mid \langle s, \tau \rangle @ t \in A \})
\]

defines a $\mathcal{INV}(B(S), \rho_F)\cdot B(\mathbb{R}_+)$-measurable map, so does

\[
s \mapsto \limsup_{t \to \infty} M(s)(\{ \tau \mid \langle s, \tau \rangle @ t \in A \}).
\]

Since these maps coincide on $Q_A$, this establishes the first part of 2.

3. Represent $X(I, A)$ for the $\rho_F$-invariant Borel set $A \subseteq X$ and the interval $I \subseteq \mathbb{R}_+$ with rational endpoints as $X(I, A) = (S \times I) \times (A \times \mathbb{R}_+) \times \text{PATHS}$, then it is clear that this a cylinder set which is $(\rho_F \times \Delta_{\mathbb{R}_+})\infty$-invariant. This establishes the second part of 2.

4. Represent for the $\rho_F$-invariant Borel sets $A_1, A_2$ and the interval $I \subseteq \mathbb{R}_+$ with rational endpoints the set $U(I, A_1, A_2)$ as in the proof of Lemma 6.2.11 as

\[
U(I, A_1, A_2) = \bigcup_{t \in Q \cap I} \left( \{ \sigma \mid \sigma @ t \in A_1 \} \cap \bigcap_{t' \in Q \cap [0, t]} \{ \sigma \mid \sigma @ t' \in A_2 \} \right),
\]

and observe that the sets involved are all invariant Borel sets, as shown in part 2 of the present proof, then part 3 follows readily. \hfill \Box

This Lemma will turn out to be instrumental in establishing our main result on bisimulations. Its proof is somewhat awkward due to the necessity of keeping track of many smooth relations at once. It indicates on the other hand that smooth equivalence relations are a versatile tool for these investigations.
6.3.2 Closure Operations

We show that $\rho_F$ coincides with a finer equivalence relation that is generated by $F$’s closure under the operations offered by the logic. It will turn out, however, that this closure is not the only one of interest: we will close $F$ also towards the future, thus, when we know that $\varphi \in F$, then we also know that $\mathcal{X}^t \varphi$ will be a member of $F$. The reason for this closure under the DP-condition will become apparent soon.

**Proposition 6.3.6** Let $F \neq \emptyset$ be a set of formulas, denote by $\rho_F$ the equivalence relation on the set of states imposed by $F$, and let $\text{wrap}(F)$ be the closure of $F$ under the logic’s operators. Then $\rho_F = \rho_{\text{wrap}(F)}$ holds, provided $F$ satisfies the DP-condition.

**Proof**

1. Because $\rho_{\text{wrap}(F)} \subseteq \rho_F$ is trivial, we need to establish the other inclusion, and since $\rho_{\text{wrap}(F)}$ is determined by the countable set $\{[\varphi] \mid \varphi \in \text{wrap}(F)\}$ of Borel sets, it is sufficient to show that $[\varphi] \in \mathcal{INV}(\mathcal{B}(S), \rho_F)$ for each $\varphi \in \text{wrap}(F)$.

2. Since for each $\varphi \in F$ we have trivially $[\varphi] \in \mathcal{INV}(\mathcal{B}(X), \rho_F)$, an inductive reasoning with Lemma 6.3.5 on the structure of $F$-state formulas and of $F$-path formulas establishes the assertion. ⊣

As an interesting direct and first consequence of Proposition 6.3.6 we obtain that the equivalence of states on the atomic propositions determines their equivalence of all formulas, provided the DP-condition is satisfied. If it is not, we force it: Define for a set $F$ of formulas

$$
dp(F) := \bigcap \{G \subseteq \mathcal{L}_P \mid F \subseteq G, G \text{ has the DP-condition}\}
$$

as the smallest set of formulas that satisfy the DP-condition (this construction is sensible because the set $\mathcal{L}_P$ of all formulas satisfies the condition under consideration).

We obtain from Proposition 6.3.6 right away:

**Corollary 6.3.7** $\rho_{\text{dp}(F)} = \rho_{\mathcal{L}_P}$.

This result is not yet fully satisfying; in practice it means that one has to have a look at the formulas in DP-closure for concluding whether or not a given property holds for all formulas. It is, however, desirable to restrict oneself to observing properties on the atomic propositions alone, and then to say that this property holds for the entirety of formulas. This is what we investigate now. The basic idea is to find a suitable representation for $\text{dp}(F)$ and then to capitalize on Corollary 5.1.10 for identifying the equivalence relation as $\rho_{\text{dp}(F)}$.

Let $F$ be a non-empty set of formulas. Define for $\Psi \subseteq \mathcal{L}_P$ the set valued map

$$
H(\Psi) := F \cup \{ \bigwedge_{1 \leq i \leq n} \varphi_i \mid n \in \mathbb{N}, \varphi_1, \ldots, \varphi_n \in \Psi\} \cup \{P_{\text{dp}}(X^{[a,b]} \varphi) \mid \varphi \in \Psi, a, b, p \text{ rational}\},
$$

then the least fixed point

$$
H_* := \mu \Psi. H(\Psi)
$$

exists by the celebrated Kleene-Knaster-Tarski Fixed Point Theorem, and

$$
H_* = \bigcup_{n \in \mathbb{N}} H^{(n)}(\emptyset)
$$
holds, with \( H^{(n)} \) as the \( n \)th iterate of \( H \). Similarly, define for a family \( A \) of Borel sets in \( X \)
\[
h(A) := \{ \langle \varphi \rangle | \varphi \in F \} \cup A \cup \{ P_{\text{dp}}(X([a,b], A)) | A \in A, a, b, p \text{ rational} \}.
\]
Again invoking the Kleene-Knaster-Tarski Theorem, we know that the smallest fixed point
\[
C_* := \mu A . h(A)
\]
exists, and can be computed through
\[
C_* = \bigcup_{n \in \mathbb{N}} h^{(n)}(\emptyset).
\]
Here \( h^{(n)} \) is of course the \( n \)th iterate of \( h \).
As witnessed by the use of the path quantifier, both constructs are closely related:

**Lemma 6.3.8** Construct the set \( H_* \) of formulas and the family \( C_* \) of Borel sets as above. Then

1. \( H_* = \text{dp}_F(F) \),
2. the \( \rho_{\text{dp}_F}(F) \)-invariant sets are generated from \( H_* \), viz., \( \sigma(C_*) = \text{INV}_F(B(X), \rho_{\text{dp}_F}(F)) \).

**Proof** 1. It is clear from the definition of the map \( H \) that \( \mu \Psi . H(\Psi) \) satisfies the DP-condition, and it is equally clear that each set of formulas that satisfies this condition and contains \( F \) contains also \( H^{(n)}(\emptyset) \) for each \( n \in \mathbb{N} \).

2. If \( \varphi \in H_* \), then the representation of \( \mu \Psi . H(\Psi) \) shows that \( \langle \varphi \rangle \in C_* \). On the other hand, it is not difficult to see that
\[
h(\langle \varphi \rangle | \varphi \in \Psi) \subseteq \{ \langle \varphi \rangle | \varphi \in H(\Psi) \} \subseteq \sigma(\{ \langle \varphi \rangle | \varphi \in H_* \}).
\]
This shows that
\[
\sigma(\{ \langle \varphi \rangle | \varphi \in H_* \}) = \sigma(C_*)
\]
establishing the second claim. \( \Box \)

Looking at the maps — which yield equivalent representations for the \( \text{dp}() \)-closure — it is noticeable that \( H \) as the version catering for formulas takes the conjunction into account, while its set-theoretic cousin \( h \) does not. This is due to the observation that the \( \sigma \)-algebra of invariant Borel sets uniquely determines the equivalence relation by Corollary 5.1.10, but that this \( \sigma \)-algebra can have many different generators which may or may not be closed with respect to finite intersection.

### 6.3.3 \( F \)-Bisimulations

Let us define \( F \)-bisimulations in order to put these results into the proper context. Define for \( F \subseteq \Sigma_F \) and for each state \( s \in S \) the set
\[
L_F(s) := \{ \varphi \in F | s \models \varphi \}
\]
as the set of all formulas in \( F \) that are satisfied by \( s \).
**Definition 6.3.9** Let $F$ be a set of formulas, then a smooth equivalence relation $\equiv_F$ is called an $F$-bisimulation iff

1. $L_F(s) = L_F(s')$, whenever $s \equiv_F s'$.

2. $K(s)(D) = K(s')(D)$, whenever $s \equiv_F s'$ and $D \in INV (B(S), \equiv_F)$.

An $F$-bisimulation is focussed on the behavior that manifests itself on the states, rather than on paths. Hence we use for its formulation the relation $K$ rather than $M$. If $\equiv_F$ is an $F$-bisimulation, condition 2 tells us that this relation is in particular a congruence (see Definition 5.2.1), so we may define the factor relation

$$K_{\equiv_F}([s]_{\equiv_F})(D) := K(s)((\eta_{\equiv_F}^{-1}[D]))$$

whenever $D \in B(S/\equiv_F)$ in a Borel set in the factor space. It has the additional property that the map $L_F : S \rightarrow F$ is constant on the equivalence classes. This observation yields a characterization of $F$-bisimulations in terms of congruences:

**Proposition 6.3.10** The following statements are equivalent for a smooth equivalence relation $\rho$ on $S$

1. $\rho$ is an $F$-bisimulation.

2. $\rho$ is a congruence for $K$ with $s \rho s' \Rightarrow L_F(s) = L_F(s')$.

Consequently, $F$-bisimilar states accept exactly the same formulas in $F$, and they behave in exactly the same way on the $\equiv_F$-invariant Borel sets. As a first result towards relating the results obtained so far to $F$-bisimulations, we see that under the mild condition of $F$ being closed under conjunctions, $\rho_F$ is actually one:

**Proposition 6.3.11** The relation $\rho_F$ is an $F$-bisimulation for each $F \subseteq L_P$, provided $F$ is closed under conjunctions.

**Proof** The definition of $\rho_F$ guarantees that $L_F(x) = L_F(x')$ is true whenever $s \rho_F s'$. Thus we need to show that $K(s)(D) = K(s')(D)$ for $s \rho_F s'$ and for each $D \in INV (B(S), \rho_F)$ holds. Define for fixed states $s, s'$ that are $\rho_F$-related

$$\mathcal{D} := \{D \in INV (B(S), \rho_F) \mid K(s)(D) = K(s')(D)\}.$$

Then $\mathcal{D}$ is a $\sigma$-algebra, and it will be enough to show that a generator of $INV (B(S), \rho_F)$ that is closed under finite intersection is contained in $\mathcal{D}$. We know from Lemma 6.3.3 that $K(s)([\varphi]) = K(s'(\varphi])$ holds for each $\varphi \in F$. But this implies with the $\pi$-$\lambda$-Theorem A.1.1 the following chain:

$$INV (B(S), \rho_F) = \sigma (\{[\varphi] \mid \varphi \in F\}) \subseteq \sigma(\mathcal{D}) \subseteq \mathcal{D} \subseteq INV (B(S), \rho_F),$$

establishing the assertion. $\dashv$

The relation $\rho_F$ is provided naturally with $F$, so it plays a prominent role among all the $F$-bisimulations (there are other $F$-bisimulations, e.g., the identity is one, but probably not the most interesting among all the candidates):
Definition 6.3.12 The states \( s, s' \in S \) are called \( F \)-bisimilar iff \( \rho_F \) \( s \) \( s' \) holds.

This is a characterization of \( F \)-bisimilarity:

Theorem 6.3.13 Let \( \emptyset \neq F \subseteq \mathcal{L}_P \) be a set of formulas which satisfy the DP-condition, then two states are \( F \)-bisimilar iff they satisfy exactly the same formulas in \( \text{wrap}(F) \).

Proof This follows immediately from Proposition 6.3.6 in conjunction with Proposition 6.3.11. \( \dashv \)

Specializing to the set of atomic formulas, we obtain at once:

Corollary 6.3.14 Two states are \( \text{dp}(P) \)-bisimilar iff they satisfy exactly the same formulas in \( \mathcal{L}_P \).

This is not yet fully satisfying for practical purposes, because one has to construct the closure \( \text{dp}(P) \) of the set of all atomic propositions, which may be done iteratively through the computation of a fixed point, as the discussion leading to Lemma 6.3.8 shows. Nevertheless it leads to an infinite process, handling a countable set of objects. But suppose we are in the situation in which both the state transitions \( K \) and the jump times \( L \) are determined through a rate function \( R \) (cp. Observation 6.2.2). Now

\[
s \models \mathcal{P}_{\text{dp}}(X^I \varphi) \iff (L(s)(I) \cdot K(s)([\varphi])) \models p
\]

as an easy computation reveals. Thus the \( \sigma \)-algebra of \( \rho_{\text{dp}(P)} \)-invariant Borel sets is determined by the \( \rho_P \)-invariant Borel sets and by the smallest \( \sigma \)-algebra \( \mathcal{I}_R \) on \( S \) that renders the map \( s \mapsto R(s)(A) \) measurable for each \( A \in \mathcal{I}_N(B(S), \rho_P) \). This is so by Lemma 6.3.8. This observation yields

Corollary 6.3.15 If \( s \mapsto R(s)(A) \) is a \( \mathcal{I}_N(B(S), \rho_P)^{B(\mathbb{R}_+)} \)-measurable map for each \( \rho_P \)-invariant Borel set \( A \), then the following conditions are equivalent for any two states \( s, s' \in X \):

1. \( s \) and \( s' \) are \( P \)-bisimilar.
2. \( s \) and \( s' \) satisfy exactly the same formulas in \( \mathcal{L}_P \).

Proof The condition implies that

\[
\mathcal{I}_N(B(X), \rho_P) = \mathcal{I}_N(B(X), \rho_{\text{dp}(P)}) = \mathcal{I}_N(B(X), \rho_{\mathcal{L}_P^P}),
\]

because all sets that are added when constructing \( \mathcal{I}_N(B(X), \rho_{\text{dp}(P)}) \) through the process described in Lemma 6.3.8 are \( \mathcal{I}_N(B(X), \rho_P) \) - measurable. Thus we infer from Corollary 5.1.10 the first equality. The second equality comes from Corollary 6.3.14. Given this equality, the assertion follows from Theorem 6.3.13. \( \dashv \)

The proof capitalizes on the uniqueness of the invariant sets for a smooth equivalence relation: since we are able to identify these sets, we may conclude what shape the relation has. This shows that a closer inspection of the invariant Borel sets bears some — probably unexpected — fruits. The condition imposed in Corollary 6.3.15 above is satisfied in the finite case whenever the rate function is constant on the equivalence classes for \( \equiv_P \). This can be checked quite efficiently once the classes are computed.

Bisimilarity for stochastic relations was discussed extensively in particular in Chapter 5, so the question of relating \( F \)-bisimilarity to that more general notion arises. A first step towards a characterization is
Proposition 6.3.16 Let $\emptyset \neq F \subseteq \mathcal{L}_F$ be a set of formulas which satisfy the DP-condition, then there exists a smooth 2-bisimulation $N : \rho_F \sim (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty$ for $M : S \sim (\mathbb{R}_+ \times S)^\infty$.

Proof From Proposition 6.3.4 we know that $(\rho_F, (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty)$ is a congruence for $M$, because $F$ satisfies the DP-condition. Thus the assertion follows from Proposition 5.4.2.

The DP-condition turns out to be crucial as a necessary condition for the two notions of bisimilarity to be related. It can said actually a bit more. We introduce for this the extension of $F$,

$$\text{ext} (F) := \{ \varphi \mid \llbracket \varphi \rrbracket \in \text{INV} (B(S), \rho_F) \}.$$  

Thus $\varphi \in \text{ext} (F)$ iff $\llbracket \varphi \rrbracket$ is $\rho_F$-invariant, so it is immediate that $F \subseteq \text{ext} (F)$, and that $\text{INV} (B(S), \rho_F) = \text{INV} (B(S), \rho_{\text{ext}(F)})$. The reason for introducing the extension is quite obviously of a strategic nature: we cannot lay our hands on $F$ directly, but we can determine whether or not a formula is in $\text{ext} (F)$ by having a look at the invariant Borel sets.

Proposition 6.3.17 The following conditions are equivalent for a set $\emptyset \neq F \subseteq \mathcal{L}_F$ of formulas:

1. $\text{ext} (F)$ satisfies the DP-condition.
2. There exists a smooth 2-bisimulation $N : \rho_F \sim (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty$ for $M : S \sim (\mathbb{R}_+ \times S)^\infty$.
3. $(\rho_F, (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty)$ is a congruence for $M$.

Proof 1. We know from Proposition 5.4.2 that the conditions 2 and 3 are equivalent, and we know from

$$\text{INV} (B(S), \rho_F) = \text{INV} (B(S), \rho_{\text{ext}(F)})$$

that $\rho_F = \rho_{\text{ext}(F)}$ by Corollary 5.1.10. Thus 1 $\Rightarrow$ 2 is just Proposition 6.3.16.

2. 2 $\Rightarrow$ 1 Abbreviate as above $\rho_F^{(\infty)} := (\Delta_{\mathbb{R}_+} \times \rho_F)^\infty$. We claim that the set

$$A_s := \{ \tau \mid \langle s, \tau \rangle \in X^I \varphi \}$$

is $\rho_F^{(\infty)}$-invariant, whenever $[\varphi] \in \text{INV} (B(S), \rho_F)$ and $s \in S$. In fact, let $\langle s, \tau \rangle \in [X^I \varphi]$ and assume that $\tau \rho_F^{(\infty)} \tau'$. By definition, $\langle s, \tau \rangle [1] = \varphi$ and $\delta(\langle s, \tau \rangle, 0) \in I$. But $\langle s, \tau \rangle [1] \rho_F \langle s, \tau' \rangle [1]$ and $\delta(\langle s, \tau \rangle, 0) = \delta(\langle s, \tau' \rangle, 0)$, so that $\langle s, \tau' \rangle [1] = \varphi$ and $\delta(\langle s, \tau' \rangle, 0) \in I$. Consequently, $\langle s, \tau' \rangle \in [X^I \varphi]$, so that $\tau' \in A_s$. Note that $A_s$ does not really depend on $s$ by the definition of $[X^I \varphi]$. Because $A_s \subseteq \text{INV} (B((R_+ \times S)^\infty), \rho_F^{(\infty)})$, we infer from Lemma 5.1.13, part 1, that for $\langle s, s' \rangle \in \rho_F$ the following holds (here $\pi_i : \rho_F^{(\infty)} \rightarrow (\mathbb{R}_+ \times S)^\infty$ are the corresponding projections, $i = 1, 2$):

$$M(s)(\{ \tau \mid \langle s, \tau \rangle \in X^I \varphi \}) = \begin{cases} 
N(\langle s, s' \rangle (\pi^{-1} [A_s]) \\
N(\langle s, s' \rangle (\pi^{-1} [A_s] \cap \rho_F^{(\infty)})) \\
N(\langle s, s' \rangle (\pi^{-1} [A_s] \cap \rho_F^{(\infty)})) \\
N(\langle s, s' \rangle (\pi^{-1} [A_s]) \\
M(s)(\{ \tau \mid \langle s', \tau \rangle \in X^I \varphi \}).
\end{cases}$$

211
But this means that
\[ s \models P_\varphi (\chi^I \varphi) \iff s' \models P_\varphi (\chi^I \varphi) \]
whenever \( s \rho_F s' \) and \( [\varphi] \in INV(B(S), \rho_F) \). Consequently, \( [P_\varphi (\chi^I \varphi)] \) is an \( \rho_F \)-invariant Borel set, thus \( P_\varphi (\chi^I \varphi) \in \text{ext}(F) \). ∎

6.4 Bibliographic Notes

Hennessy and Milner introduced in their 1980 paper [44] a very simple and negation free modal logic and related bisimilarity of image-finite Kripke models to the equivalence relation “accepting the same formulas” on states. Subsequently, the seminal paper by Larsen and Skou [59] introduced stochastic Kripke models, albeit over discrete state spaces, and established a Hennessy-Milner like Theorem for simple modal logics, among others a variant of the Hennessy-Milner logic (in which the diamond operator \( \Diamond \varphi \) is replaced by a family of diamond operators \( \Diamond_q \varphi \) with \( 0 \leq q \leq 1 \)). Changing the stage from discrete to analytic state spaces, Desharnais, Edalat and Panangaden investigated the problem of bisimilarity again. The research reported in [20] takes an analytic state space with universally measurable transition functions as a basic scenario. A Hennessy-Milner theorem is proved; the proof’s idea is to produce a co-span of morphisms through injections into a suitably factored sum. This idea has left its traces in various parts of the present exposition. But the situation considered here is structurally subtly different: universal measurability, as assumed in [20], requires a somewhat elaborate completion process using all finite measures on that space. This is mainly due to the fact that the existence of semi-pullbacks could only be established by Edalat under these circumstances. After the existence of semi-pullbacks could be established also for relations on the Borel sets of an analytic space (see Chapter 4 and [25]), the question of bisimulations became tractable also for the more general and natural case of analytic spaces with their Borel structure. These spaces are structurally much simpler and do not need additional considerations, since they are given through the morphisms of measurable spaces and nothing else (so one could work with them even if one would want to do without the real numbers).

In [25] a generalization of [20] is established for those labeled Markov transition systems which work over a Polish (rather than an analytic) state space and which have a certain smallness property. This technical condition is lifted in the present exposition. This is so since the technique of factoring stochastic relations is better understood now. Apart from a much wider class of modal logics which can be dealt with now (as witnessed in Section 3.5), the present discussion proposes a more general technical approach.

The logic \( pCTL^* \) is defined and studied in [15] from where our brief illustrative exposition is taken; \( CSL \) [7] is a stochastic version and variant of the popular logic \( CTL \) for model checking [18]. The logic has considerable expressive power, as is demonstrated convincingly in [7]. Recently, Desharnais and Panangaden [22] have proposed an interpretation of a subset of \( CSL \) over a continuous domain, hereby providing a general framework for the treatment of bisimulations. The originally given interpretation in [7] is based on a finite state space in order to investigates the computational side of model checking using CSL. A comparison with [22] suggests that the wide and well-assorted toolkit provided by probabilities over analytic spaces is a welcome addition for investi-
gating the properties of this logic. This is particularly true when it comes to investigating bisimulations.
Appendix A

Measure Theory and Topology — A Refresher

Contents

A.1 Measurable Spaces ................................................. 215
A.2 Polish and Analytic Spaces ........................................ 217
  A.2.1 Manipulating Polish Topologies ........................... 218
  A.2.2 Analytic Spaces .............................................. 218
  A.2.3 Measurable Selectors ......................................... 220
A.3 Probability Measures ............................................. 221
  A.3.1 Weak Topology .................................................. 222
  A.3.2 Applications of the $\pi$-$\lambda$-Theorem ................. 223
  A.3.3 Projective Systems ............................................ 224
  A.3.4 Universal Measurability ..................................... 227

This Appendix collects some results and methods from measure theory and from topology for the reader’s convenience. Most of it can be found in [42, 73, 10, 35, 45]. We will make also use of results of the theory of Borel sets which is presented in accessible form in [51, 88].

A.1 Measurable Spaces

A measurable space $(M, \mathcal{M})$ consists of a set $M$ with a $\sigma$-algebra $\mathcal{M}$, which is an algebra of subsets of $M$ that is closed under countable unions (hence countable intersections or countable disjoint unions). If $\mathcal{M}_0$ is a family of subsets of $M$, then

$$\sigma(\mathcal{M}_0) = \bigcap\{\mathcal{M} \mid \mathcal{M} \text{ is a } \sigma\text{-algebra on } M \text{ with } \mathcal{M}_0 \subseteq \mathcal{M}\}$$

is the smallest $\sigma$-algebra on $M$ which contains $\mathcal{M}_0$. This construction works since the power set $\mathcal{P}(M)$ is a $\sigma$-algebra on $M$. Take for example as a generator $\mathcal{I}$ all open intervals in the real numbers $\mathbb{R}$, then $\sigma(\mathcal{I}) =: \mathcal{B}(\mathbb{R})$ is the $\sigma$-algebra of real Borel sets. We will encounter the Borel sets again in Section A.2.

215
If \((N, \mathcal{N})\) is another measurable space, then a map \(f : M \to N\) is called \(\mathcal{M}\)-\(\mathcal{N}\)-measurable iff the inverse image under \(f\) of each set in \(\mathcal{N}\) is a member of \(\mathcal{M}\), hence iff \(f^{-1}[G] \in \mathcal{M}\) holds for all \(G \in \mathcal{N}\). Suppose that \(\mathcal{N}\) is generated by the family \(N_0\) of subsets of \(N\), so that \(\mathcal{N} = \sigma(N_0)\) holds, then \(f\) is \(\mathcal{N}\)-\(\mathcal{M}\)-measurable iff \(f^{-1}[G] \in \mathcal{M}\) holds for all \(G \in N_0\). Thus we may restrict the attention to inverse images of sets from a generator, which is sometimes easier to handle. For example, a real valued function \(f : M \to \mathbb{R}\) on \(M\) is \(\mathcal{M}\)-\(\mathcal{B}(\mathbb{R})\)-measurable iff \(\{m \in M \mid f(m) \bowtie t\} \in \mathcal{M}\) holds for each \(t \in \mathbb{R}\); the relation \(\bowtie\) may be taken from \(<, \leq, \geq, >\).

If \((M, \mathcal{M})\) is a measurable space and \(f : M \to N\) is a map, then

\[
\mathcal{N} := \{D \subseteq N \mid f^{-1}[D] \in \mathcal{M}\}
\]

is the largest \(\sigma\)-algebra \(N_0\) on \(N\) that renders \(f\) \(\mathcal{M}\)-\(N_0\)-measurable (\(\mathcal{N}\) is the final \(\sigma\)-algebra w.r.t. \(f\)). If \(g : P \to M\) is a map, then

\[
g^{-1}[\mathcal{M}] := \{g^{-1}[E] \mid E \in \mathcal{M}\}
\]

is the smallest \(\sigma\)-algebra \(\mathcal{P}_0\) on \(P\) that renders \(g : \mathcal{P}_0 \to \mathcal{M}\) measurable (accordingly, \(g^{-1}[\mathcal{M}]\) is called initial w.r.t. \(f\)).

Let \((M_i, \mathcal{M}_i)_{i \in I}\) be a family of measurable spaces, then the product-\(\sigma\)-algebra \(\bigotimes_{i \in I} \mathcal{M}_i\) denotes that smallest \(\sigma\)-algebra \(\mathcal{M}_0\) on \(\prod_{i \in I} M_i\) which makes all the projections

\[
\pi_j : (m_i \mid i \in I) \mapsto m_j
\]

a \(\mathcal{M}_0\)-\(\mathcal{M}_j\)-measurable map for each index \(j \in I\). It is not difficult to see that

\[
\bigotimes_{i \in I} \mathcal{M}_i = \sigma(\big\{\prod_{i \in I} E_i \mid \forall i \in I : E_i \in \mathcal{M}_i, E_i = M_i \text{ for all but a finite number of indices}\})
\]

The generating sets are sometimes called cylinder sets; we make frequent use of them. For \(I = \{1, 2\}\), the \(\sigma\)-algebra \(\mathcal{M}_1 \otimes \mathcal{M}_2\) is generated from the set of measurable rectangles

\[
\{E_1 \times E_2 \mid E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2\}
\]

Dually, the sum \((X_1 + X_2, \mathcal{A}_1 + \mathcal{A}_2)\) of the measurable spaces \((X_1, \mathcal{A}_1)\) and \((X_2, \mathcal{A}_2)\) is defined through the largest \(\sigma\)-algebra \(\mathcal{D}\) on the sum \(X_1 + X_2\) that makes both injections \(\mathcal{A}_1\)-\(\mathcal{D}\) and \(\mathcal{A}_2\)-\(\mathcal{D}\)-measurable.

An important tool is the \(\pi\)-\(\lambda\)-Theorem which makes it sometimes simpler to identify the \(\sigma\)-algebra generated from some family of sets, see [51, Theorem 10.1, iii]):

**Theorem A.1.1 (\(\pi\)-\(\lambda\)-Theorem)** Let \(\mathcal{P}\) be a family of subsets of a set \(X\) that is closed under finite intersections (a \(\pi\)-class). Then \(\sigma(\mathcal{P})\) is the smallest \(\lambda\)-class containing \(\mathcal{P}\), where a family \(\mathcal{L}\) of subsets of \(X\) is called a \(\lambda\)-class iff it is closed under complements and countable disjoint unions.

We need sometimes to approximate measurable functions by linear combinations of indicator functions. Define for \(A \subseteq N\) the indicator function

\[
\chi_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}
\]
Clearly, if $\mathcal{N}$ is a $\sigma$-algebra on $N$, then $A \in \mathcal{N}$ iff $\chi_A$ is a $\mathcal{N} \cdot \mathcal{B}(\mathbb{R})$-measurable function. A measurable step function

$$f = \sum_{i=1}^{n} \alpha_i \cdot \chi_{A_i}$$

is a linear combination of indicator functions with $A_i \in \mathcal{N}$. The following statement is folklore in measure theory, it is rather helpful in many situations when we have information about the behavior of a construction for measurable sets (i.e., for indicator functions), when things behave linearly, and when it is closed under monotone convergence.

**Proposition A.1.2** Denote for a measurable space $(N, \mathcal{N})$ by

$$\mathcal{F}(N, \mathcal{N}) := \{ f : N \to \mathbb{R} \mid f \text{ is $\mathcal{N} \cdot \mathcal{B}(\mathbb{R})$ measurable and bounded} \}$$

the linear space of all bounded measurable real functions on $N$. Then

1. For $f \in \mathcal{F}(N, \mathcal{N})$ with $f \geq 0$ there exists an increasing sequence $(f_n)_{n \in \mathbb{N}}$ of step functions $f_n \in \mathcal{F}(N, \mathcal{N})$ with

$$f(x) = \sup_{n \in \mathbb{N}} f_n(x)$$

for all $x \in X$.

2. For $f \in \mathcal{F}(N, \mathcal{N})$ there exists a sequence $(f_n)_{n \in \mathbb{N}}$ of step functions $f_n \in \mathcal{F}(N, \mathcal{N})$ with

$$f(x) = \lim_{n \to \infty} f_n(x)$$

for all $x \in X$.

**Convention.** Measurability of real-valued functions always means measurability with respect to the Borel sets $\mathcal{B}(\mathbb{R})$ of the real numbers, unless otherwise stated.

### A.2 Polish and Analytic Spaces

General measurable spaces are sometimes too general for supporting specific structures. We deal with Polish and analytic spaces which are general enough to support interesting applications but have specific properties which make life easier from a measure theoretic point of view. A Polish space $X$ is a topological space the topology of which is metrizable through a complete metric, and which has a countable dense subset. The Borel sets $\mathcal{B}(X)$ of $X$ form the smallest $\sigma$-algebra on $X$ that contains the open subsets of $X$; if $D$ is a dense subset, then it is known that $\mathcal{B}(X) = \sigma(\{ B_r(x) \mid x \in D, r \in \mathbb{Q} \})$, where $B_r(x) := \{ x' \in X \mid d(x, x') < r \}$ is the open ball around $x$ with radius $r$ in a metric $d$ inducing the topology. This representation implies that $\mathcal{B}(X)$ is countably generated. When writing down a Polish space $X$, we always have the measurable space $(X, \mathcal{B}(X))$ in mind; we omit the notation of the $\sigma$-algebra in order to avoid further burdening the notation. This applies in particular to the real numbers: we always assume that the real numbers $\mathbb{R}$ (or the non-negative real numbers $\mathbb{R}_+$) are equipped with the Borel sets $\mathcal{B}(\mathbb{R})$ (resp. $\mathcal{B}(\mathbb{R}_+)$). When talking about Borel measurable maps between Polish spaces,
we have measurability with respect to the Borel σ-algebras in mind, unless specified otherwise. Polish spaces have a number of pleasant properties that are being made use of here. One example is the closedness under countable Cartesian products: If \((X_n)_{n \in \mathbb{N}}\) is a countable family of Polish spaces, then their Cartesian product is a Polish space, and
\[
\mathcal{B}(\prod_{n \in \mathbb{N}} X_n) = \bigotimes_{n \in \mathbb{N}} \mathcal{B}(X_n)
\]
holds (this applies of course to the finite case as well). This observation yields a practical representation of the Borel sets in a countable product through cylinder sets. Similarly, Polish spaces are closed under countable sums. A closed subspace of a Polish space is Polish again, so is an open subspace. More general: if \(X\) is a Polish space, and if \(A \subseteq X\) can be represented as \(A = \bigcap_{n \in \mathbb{N}} G_n\), where \(G_n \subseteq X\) is open, then \(A\) is a Polish space in its own right (then \(A\) is called a \(G_\delta\)-set; each open and each closed set is a \(G_\delta\)-set, as we know from the theory of Polish spaces). Conversely, each Polish space can be represented as a \(G_\delta\)-set in the hypercube \([0,1]^\mathbb{N}\); this is the famous characterization of Polish spaces due to Alexandrov [56, III.33.VI], which is used e.g. in section 4.2.2.

When the context is clear, we write down topological or measurable spaces without their topologies and σ-algebras, resp., and the Borel sets are always understood with respect to the topology under consideration.

**A.2.1 Manipulating Polish Topologies**

The topology determines the Borel sets of a Polish space, but this relation is somewhat flexible: given a measurable map between Polish spaces, we can find a finer Polish topology on the domain, which has the same Borel sets, and which renders the map continuous, see [88, Corollary 3.2.5].

**Proposition A.2.1** Let \((X, T)\) be a Polish space, \((Y, T')\) be a second countable metric space. If \(f : X \to Y\) is a Borel measurable map, then there exists a Polish topology \(T_0\) on \(X\) such that \(T_0\) is finer than \(T\) (hence \(T \subseteq T_0\)), \(T\) and \(T_0\) have the same Borel sets, and \(f\) is \(T_0 - T'\) continuous. \(\square\)

This property is most useful, because it permits rendering measurable maps continuous, when they go into a second countable metric space (thus in particular into a Polish space). On the level of Borel sets, Proposition A.2.1 has the following counterpart, which is proved in [88, Corollary 3.2.6]:

**Proposition A.2.2** Let \((X, T)\) be a Polish space. If \((B_n)_{n \in \mathbb{N}}\) is a sequence of Borel sets in \(X\), then there exists a Polish topology \(T_0\) on \(X\) such that \(T_0\) is finer than \(T\), \(T\) and \(T_0\) have the same Borel sets, and each \(B_n\) is clopen (i.e., closed and open) in \(T_0\).

**A.2.2 Analytic Spaces**

An analytic space \(X\) is a measurable space \((X, A)\) that is Borel isomorphic to the image of a Polish space under a Borel measurable map. Hence we have the following picture: \(P \xrightarrow{f} f[P] \xrightarrow{h} X\) where \(P\) and \(Q\) are Polish spaces, \(f : P \to Q\) is a Borel measurable map.
A.2 Polish and Analytic Spaces

map, \( h : f[P] \to X \) is a Borel isomorphism, where \( f[P] \) is endowed with the trace \( \sigma \)-algebra \( \{W \cap f[P] \mid W \in B(Q)\} \) of the Borel sets on \( f[P] \), and \( X \) is endowed with \( \mathcal{A} \). The \( \sigma \)-algebra on an analytic space \( X \) is usually also referred to as the Borel sets on \( X \) and written down as \( B(X) \). There are some equivalent characterizations of analytic spaces, the reader is referred to [88, 51], and for a concise general discussion of factoring analytic spaces with countably generated (smooth) equivalence relations to [5].

Call a measurable space \((X, \mathcal{A})\) separable iff the \( \sigma \)-algebra \( \mathcal{A} \) has a countable set \( (A_n)_{n \in \mathbb{N}} \) of generators which separates points, i.e. given \( x, x' \in X \) with \( x \neq x' \) there exists \( A_n \) which contains exactly one of them. A Polish space is separable as a measurable space, so is an analytic space.

Sometimes second countable metric spaces are called separable, and this analogy is justified:

**Lemma A.2.3** For a separable measurable space \((X, \mathcal{A})\) there exists a second countable metric topology \( T \) on \( X \) such that \( B(X, T) = \mathcal{A} \).

**Proof** The assertion follows from [88, Proposition 3.3.2 and Remark 3.3.3]. ⊢

This innocently looking statement has some remarkable consequences for our context. Just as an appetizer:

**Corollary A.2.4** Let \((Y, B)\) be a separable measurable space. Then

1. The diagonal is measurable in the product, i.e.,
   \[
   \Delta_{Y \times Y} := \{(y, y) \mid y \in Y\} \in B \otimes B.
   \]
2. If \( f_i : X_i \to Y \) is \( \mathcal{A}_i \)-measurable, where \((X_i, \mathcal{A}_i)\) is a measurable space \((i = 1, 2)\), then
   \[
   f_1^{-1}[B] \otimes f_2^{-1}[B] = (f_1 \times f_2)^{-1}[B \otimes B].
   \]

**Proof** 0. The proofs are based on standard arguments; they are given for the reader's convenience. Assertion ?? follows from Lemma A.3.10.

1. It is well-known [73, Theorem I.3.3] that property 1 holds for second countable metric spaces, so by Lemma A.2.3 it holds for separable metric spaces as well.
2. The product \( \sigma \)-algebra \( B \otimes B \) is generated by the rectangles \( B_1 \times B_2 \) with \( B_i \) taken from some generator \( B_0 \) for \( B \) \((i = 1, 2)\). Since
   \[
   (f_1 \times f_2)^{-1}[B_1 \times B_2] = f_1^{-1}[B_1] \times f_2^{-1}[B_2],
   \]
   we see that
   \[
   (f_1 \times f_2)^{-1}[B \otimes B] \subseteq f_1^{-1}[B] \otimes f_2^{-1}[B].
   \]

This is true without the assumption of separability. Now let \( T \) be a second countable metric topology on \( Y \) with \( B = B(Y, T) \) and let \( T_0 \) be a countable base for the topology. Then
   \[
   T_p := \{T_1 \times T_2 \mid T_1, T_2 \in T_0\}
   \]
is a countable base for the product topology \( T \otimes T \), and (this is the crucial property)
   \[
   B \otimes B = B(Y \times Y, T \otimes T).
   \]
holds \[88, \text{Proposition 3.1.23}\]. Since for \( T_1, T_2 \in T_0 \) clearly
\[
f_1^{-1} [T_1] \times f_2^{-1} [T_2] \in (f_1 \times f_2)^{-1} [T] \subseteq (f_1 \times f_2)^{-1} [B \otimes B]
\]
holds, the non-trivial inclusion is inferred from the fact that the smallest \( \sigma \)-algebra containing \( \{ f_1^{-1} [T_1] \times f_2^{-1} [T_2] \mid T_1, T_2 \in T_0 \} \) equals \( f_1^{-1} [B] \otimes f_2^{-1} [B] \).

Sometimes one starts with a measurable space rather than with a topological one: A Standard Borel space \((X, \mathcal{A})\) is a measurable space such that the \( \sigma \)-algebra \( \mathcal{A} \) equals \( \mathcal{B}(X, T) \) for some Polish topology \( T \) on \( X \).

We need two important results from the theory of Borel sets: Souslin’s Theorem gives a criterion for an analytic set to be Borel, and the Blackwell-Mackey-Theorem on countably generated sub-\( \sigma \)-algebras of the Borel sets of an analytic space.

Let \( X \) be an analytic space, then a subset \( A \subseteq X \) is called analytic iff there exists a Borel map \( f : P \rightarrow X \) for a Polish space with \( A = f [P] \). The complement of an analytic set is called co-analytic. Souslin’s famous Theorem [88, Theorem 4.4.3] yields a criterion for an analytic set to be Borel:

**Theorem A.2.5 (Souslin)** Let \( X \) be an analytic space. An analytic set that is also co-analytic is a Borel set. \( \dashv \)

The Blackwell-Mackey-Theorem [88, Theorem 4.5.7] analyzes those Borel sets that are unions of \( \mathcal{A} \)-atoms for a sub-\( \sigma \)-algebra \( \mathcal{A} \subseteq \mathcal{B}(X) \). Recall that a set \( W \in \mathcal{A} \) is an \( \mathcal{A} \)-atom (or simply an atom) iff for each \( V \in \mathcal{A} \) with \( V \subseteq W \) either \( V = \emptyset \) or \( V = W \) holds.

**Theorem A.2.6 (Blackwell-Mackey)** Let \( X \) be an analytic space and \( \mathcal{A} \subseteq \mathcal{B}(X) \) be a countably generated sub-\( \sigma \)-algebra of the Borel sets of \( X \). If \( B \subseteq X \) is a Borel set that is a union of atoms of \( \mathcal{A} \), then \( B \in \mathcal{A} \). \( \dashv \)

### A.2.3 Measurable Selectors

Assume that \( X \) and \( Z \) is a measurable resp. a Polish space. Consider a set valued map \( R : X \rightarrow \mathcal{P}(Z) \), (equivalently, a relation \( R \subseteq X \times Z \). We will not distinguish too narrowly between relations and set valued maps, so that even for a relation \( R \) the set \( R(x) \) will be defined.) If \( R(x) \) always takes closed and non-empty values, and if the weak inverse
\[
(\exists R)(G) := \{ x \in X \mid R(x) \cap G \neq \emptyset \}
\]
is a measurable set, whenever \( G \subseteq Z \) is open, then \( R \) is called a weakly measurable relation on \( Y \times Z \). Since \( Z \) is Polish, \( R \) is a measurable relation iff the strong inverse
\[
(\forall R)(F) := \{ x \in X \mid R(x) \subseteq F \}
\]
is measurable, whenever \( F \subseteq Z \) is closed [46, Theorem 3.5]. \( R \) is called \( \mathcal{C} \)-measurable iff for a compact set \( C \subseteq Z \) the weak inverse
\[
\exists F(C) := \{ x \in X \mid F(x) \cap C \neq \emptyset \}
\]
is a Borel set in \( X \). A selector \( s \) for such a relation \( R \) is a single-valued map \( s : X \rightarrow Y \) such that \( s(x) \in F(x) \) holds for each \( x \in X \).
A.3 Probability Measures

Both weak and strong inverse of a relation invoke an analogy to modal logic. Let \([\varphi]\) be the set of states in a Kripke model for which formula \(\phi\) holds, and assume that \(R\) is the relation in that model. Then it is immediate that

\[(\exists R)[[\varphi]] = [\diamond \phi]\]

and

\[(\forall R)[[\varphi]] = [\Box \phi]\]

hold.

Weakly measurable relations can be represented through measurable selectors (some times called a Castaing representation). This representation implies in particular that a weakly measurable set valued map has a measurable selector. It is established in [94, Theorem 4.2.e].

**Proposition A.2.7** Given the Polish spaces \(X\) and \(Z\) and a set valued map \(R \subseteq X \times Z\).

1. \(R\) is weakly measurable iff there exists a sequence \((f_n)_{n \in \mathbb{N}}\) of Borel measurable maps \(f_n : X \to Z\) such that \(\{f_n(x) \mid n \in \mathbb{N}\}\) is dense in \(R(x)\) for each \(x \in X\).

2. If \(R\) is \(C\)-measurable, then \(R\) has a measurable selector.

Postulating measurability for \(\exists R(C)\) for open or for closed sets \(C\) leads to the general notion of a measurable relation. These relations are a valuable tool in such diverse fields as stochastic dynamic programming [94] and descriptive set theory [51]. Overviews are provided in [88, Chapter 5] and [46, 94].

A.3 Probability Measures

A probability measure on the measurable space \((N,\mathcal{N})\) is a monotone and \(\sigma\)-additive map \(\mu : \mathcal{N} \to [0,1]\) with \(\mu(\emptyset) = 0\) and \(\mu(N) = 1\). That \(\mu\) is \(\sigma\)-additive means that

\[\mu(\bigcup_{i \in \mathbb{N}} D_i) = \sum_{i \in \mathbb{N}} \mu(D_i)\]

holds whenever \((D_n)_{n \in \mathbb{N}}\) is a countable family of mutually disjoint sets in \(\mathcal{N}\). Denote by \(\mathcal{P}(N,\mathcal{N})\) the set of all probability measures on \((N,\mathcal{N})\). Occasionally we will use sub-probability measures: they are defined like probability measures with the exception that the entire spaces is assigned a mass which does not exceed unity; \(\mathcal{S}(N,\mathcal{N})\) is the set of all sub-probability measures on \((N,\mathcal{N})\).

A rather important tool is the Monotone Convergence Theorem, which yields the analogue to \(\sigma\)-additivity for the integral:

**Proposition A.3.1** Let \(f \in \mathcal{F}(N,\mathcal{N})\) for the measurable space \((N,\mathcal{N})\) be a nonnegative and bounded measurable function with \(f \geq 0\), assume that \(0 \leq f_1 \leq f_2 \leq \ldots\) is a monotonically increasing sequence \((f_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}(N,\mathcal{N})\) with \(f = \sup_{n \in \mathbb{N}} f_n\), and let \(\mu \in \mathcal{S}(N,\mathcal{N})\) be a sub-probability measure. Then

\[\int_N f \, d\mu = \lim_{n \to \infty} \int_N f_n \, d\mu.\]
A.3.1 Weak Topology

The set $S(\mathcal{N},\mathcal{N})$ is turned into a measurable space in its own right. Consider for $E \in \mathcal{N}$ the evaluation map $\mu \mapsto \mu(E)$ that assigns each measure its value at $E$. The initial $\sigma$-algebra $\mathcal{N}^\ast$ which makes all evaluation maps measurable is called the weak-* $\sigma$-algebra. Let $X$ be a metric space, then $S(X) = S(X,\mathcal{B}(X))$ is usually equipped with the topology of weak convergence. This is the smallest topology on $S(X)$ which makes the map $\mu \mapsto \int_X f \, d\mu$ continuous for each continuous and bounded $f : X \to \mathbb{R}$. Denote by $C(X)$ the linear space of all these functions, and by $\rightarrow_w$ convergence in this topology.

This topology is characterized through the famous Portmanteau Theorem [73, Theorem II.6.1]:

Proposition A.3.2 The following conditions are equivalent for a sequence $(\mu_n)_{n \in \mathbb{N}}$ and a measure $\mu \in S(X)$ for a Polish space $X$:

1. $\mu_n \rightarrow_w \mu$,
2. $\int_X f \, d\mu_n \rightarrow \int_X f \, d\mu$ for each bounded and continuous $f : X \to \mathbb{R}$,
3. $\liminf_{n \to \infty} \mu_n(F) \leq \mu(F)$ for each closed subset $F \subseteq X$.

It is well known that for second countable $X$ this topology is also second countable, that the discrete measures are dense, and that $X$ Polish implies that $S(X)$ is also Polish [73, Theorems II.6.3, II.6.5]. To be specific, let $d$ be the metric on $X$, and define

$$d(x, A) := \inf \{d(x, y) \mid y \in X\}$$

as the distance of $x \in X$ to the subset $A \subseteq X$, then the Prohorov metric $d_P$ on $S(X)$ is defined through

$$d_P(\tau_1, \tau_2) := \inf \{\varepsilon > 0 \mid \forall A \in \mathcal{B}(X) : \tau_1(A) \leq \tau_2(A^\varepsilon) + \varepsilon \wedge \tau_1(A) \leq \tau_2(A^\varepsilon) + \varepsilon\}$$

with $A^\varepsilon := \{y \in X \mid d(y, A) < \varepsilon\}$ as the set of all elements of $X$ having distance less that $\varepsilon$ from $A$. Then this metric metrizes the topology of weak convergence, see [10, Theorem 6.8]. $X$ is homeomorphic to the subset $\{\delta_x \mid x \in X\}$ of Dirac measures.

A subset $C \subseteq S(X)$ of the sub-probability measures of a Polish space $X$ is relatively compact iff it is uniformly tight [73, Theorem II.6.7]:

Proposition A.3.3 Let $C \subseteq S(X)$ for Polish $X$. Then the following conditions are equivalent:

1. The closure of $C$ in the weak topology is compact.
2. Given $\varepsilon > 0$ there exists a compact subset $K \subseteq X$ such that $\sup_{\mu \in C} \mu(X \setminus K) < \varepsilon$.

Because a single measure forms a compact set, we obtain as a Corollary that a measure on a Polish space is tight (actually, the argumentation goes the other way around, but this order is simpler when quoting the results):
Corollary A.3.4 Given $\mu \in \mathcal{S}(X)$ for Polish $X$ and $\epsilon > 0$ there exists a compact subset $C \subseteq X$ with $\mu(X \setminus C) < \epsilon$. $\dagger$

The $\sigma$-algebra of Borel sets for the topology of weak convergence is just the weak*-\-$\sigma$-algebra [51, Theorem 17.24].

A.3.2 Applications of the $\pi$-$\lambda$-Theorem

The $\pi$-$\lambda$-Theorem is used mainly when exploring measure extensions. Suppose that $\mu_n$ is a probability measure on the measurable space $(X_n, A_n)$ for each $n \in \mathbb{N}$, and define for a cylinder set

$$\hat{\mu} \left( \prod_{n \in \mathbb{N}} A_n \right) := \prod_{n \in \mathbb{N}} \mu_n(A_n).$$

Observe that in this infinite product all but a finite number of factors equal unity. Then $\hat{\mu}$ extends to a probability measure $\mu^#$ on $(\prod_{n \in \mathbb{N}} A_n, \bigotimes_{n \in \mathbb{N}} A_n)$; in particular,

$$\mu^#(A_1 \times \ldots A_n \times \prod_{j > n} X_j) = \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n)$$

holds. Accordingly, $\mu^#$ is called the product measure of $(\mu_n)_{n \in \mathbb{N}}$ and denoted by $\bigotimes_{n \in \mathbb{N}} \mu_n$.

Of course, a finite product is also available. The $\pi$-$\lambda$-Theorem assures us that the extension is unique.

Another application is given when applying horizontal or vertical cuts from a measurable set in a product and then asking about measurability of associated maps. Let $(X, A)$ and $(Y, B)$ be measurable spaces, and define for $D \in A \otimes B$ the vertical cut

$$D_x := \{ y \in Y \mid \langle x, y \rangle \in D \}$$

and the horizontal cut

$$D^y := \{ x \in X \mid \langle x, y \rangle \in D \}.$$

Lemma A.3.5 Let $(X, A)$ and $(Y, B)$ be measurable spaces, and fix $D \in A \otimes B$. The map $\langle \nu, x \rangle \mapsto \nu(D_x)$ is a $B^* \otimes A$-measurable map on $\mathcal{S}(Y, B) \times X$.

Proof Consider

$$\mathcal{D} := \{ D \in A \otimes B \mid \langle \nu, x \rangle \mapsto \nu(D_x) \text{ is } B^* \otimes A \text{- measurable} \}.$$ 

Since $((X \times Y) \setminus D)_x = Y \setminus (D_x)$ and $\bigcup_{n \in \mathbb{N}} D_n)_x = \bigcup_{n \in \mathbb{N}} (D_n)_x$, it is clear that $\mathcal{D}$ is closed under taking complements and countable disjoint unions. Now let $D = A \times B$ with $A \in A, B \in B$. Then $\nu(D_x) = \chi_A(x) \cdot \nu(B)$, thus $\langle \nu, x \rangle \mapsto \nu(D_x)$ is $B^* \otimes A$-measurable, because it may be composed as a product, hence in a measurable way from projections onto $\mathcal{S}(Y, B)$ resp. $X$. But this implies that all measurable rectangles are members of $\mathcal{D}$, and since the set of all these rectangles is closed under finite intersections, $\mathcal{D}$ equals the $\sigma$-algebra generated from them, which coincides with $A \otimes B$. The assertion is hence true for all measurable subsets of the product. $\dagger$

Lemma A.3.5 entails that both $x \mapsto \nu(D_x)$ and $\nu \mapsto \nu(D_x)$ are measurable (but the Lemma says considerably more: it establishes joint measurability).
A.3.3 Projective Systems

We need for the fixed point modelling the *while-*loop in Ludwig (section 2.4) and for the interpretation of the logic CSL (section 6.2) the projective limit of a projective family of stochastic relations. Denote by $X^\infty := \prod_{k \in \mathbb{N}} X$ the infinite product of $X$ with itself.

**Definition A.3.6** Let $X$ be a Polish space, and $(\mu_n)_{n \in \mathbb{N}}$ a sequence of probability measures $\mu_n \in \mathcal{P}(X^n)$. This sequence is called a projective system iff

$$\mu_n(A) = \mu_{n+1}(A \times X)$$

for all $n \in \mathbb{N}$ and all Borel sets $A \in \mathcal{B}(X^n)$. A probability measure $\mu_\infty \in \mathcal{P}(X^\infty)$ is called the projective limit of the projective system $(\mu_n)_{n \in \mathbb{N}}$ iff

$$\mu_n(A) = \mu_\infty(A \times \prod_{j>n} X)$$

for all $n \in \mathbb{N}$ and $A \in \mathcal{B}(X^n)$.

Thus a sequence of measures is a projective system iff each measure is the projection of the next one, its projective limit is characterized through the property that its values on cylinder sets coincides with the value of a member of the sequence, after taking projections.

It is not immediately obvious that a projective limit exists. The basic idea is to define the limit on the cylinder sets and then to extend this premeasure — but it has to be established that it is indeed a premeasure. The crucial property is that $\mu_{nk}(A_k) \to 0$ whenever $(A_n)_{n \in \mathbb{N}}$ is a sequence of cylinder sets $A_k$ (with at most $n_k$ components that do not equal $X$) that decreases to $\emptyset$. This property is difficult to establish without topological assumptions (this is why we have postulated that the base space $X$ is Polish).

The central statement is

**Proposition A.3.7** Let $X$ be a compact metric space. Then a unique projective limit $\mu_\infty$ exists for the projective system $(\mu_n)_{n \in \mathbb{N}}$.

**Proof** 1. Let $A = A'_k \times \prod_{j>k} X$ be a cylinder set with $A'_k \in \mathcal{B}(X^k)$, then define $\mu^*(A) := \mu_k(A'_k)$, then $\mu^*$ is well defined, since the sequence forms a projective system. In order to show that $\mu^*$ is a premeasure on the cylinder sets, we have to take a decreasing sequence $(A_n)_{n \in \mathbb{N}}$ of cylinder sets with $\bigcap_{n \in \mathbb{N}} A_n = \emptyset$ and show that $\inf_{n \in \mathbb{N}} \mu^*(A_n) = 0$. In fact, suppose that $(A_n)_{n \in \mathbb{N}}$ is decreasing with $\mu^*(A_n) \geq \delta$ for all $n \in \mathbb{N}$, then we show that $\bigcap_{n \in \mathbb{N}} A_n \neq \emptyset$.

We can write

$$A_n = A'_n \times \prod_{j>k_n} X$$

for some $A'_n \in \mathcal{B}(X^{k_n})$. From Corollary A.3.4 we get for each $n$ a compact set $K'_n \subseteq A'_n$ such that $\mu_{k_n}(A'_n \setminus K'_n) < \delta/4^{-n}$. Because $X^\infty$ is compact by Tichonov’s Theorem,

$$K''_n := K'_n \times \prod_{j>k_n} X$$
is a compact set, and $K_n := \bigcap_{j=1}^{n} K_j'' \subseteq A_n$ is compact as well, with

$$
\mu^*(A_n \setminus K_n) \leq \mu^*\left(\bigcup_{j=1}^{n} A_n'' \setminus K_j''\right)
$$

$$
\leq \sum_{j=i}^{n} \mu^*(A_j'' \setminus K_j'')
$$

$$
= \sum_{j=1}^{n} \mu_{k_j}(A_j' \setminus K_j')
$$

$$
\leq \sum_{j=1}^{\infty} \delta \cdot 4^{-j}
$$

$$
= \delta/3.
$$

Thus $(K_n)_{n \in \mathbb{N}}$ is a decreasing sequence of non-empty compact sets, consequently,

$$
\emptyset \neq \bigcap_{n \in \mathbb{N}} K_n \subseteq \bigcap_{n \in \mathbb{N}} A_n.
$$

2. Since the cylinder sets generate the Borel sets of $X^\infty$, and since $\mu^*$ is a premeasure, we know that there exists a unique extension $\mu_\infty \in \mathfrak{P}(X^\infty)$ to it. Clearly, if $A \subseteq X^n$ is a Borel set, then

$$
\mu_\infty(A \times \prod_{j>n} X) = \mu^*(A \times \prod_{j>n} X) = \mu_n(A),
$$

so we have constructed a projective limit.

3. Suppose that $\mu'$ is another probability measure in $\mathfrak{P}(X^\infty)$ that has the desired property. Consider

$$
\mathcal{D} := \{D \in \mathcal{B}(X^\infty) \mid \mu_\infty(D) = \mu'(D)\}.
$$

It is clear the $\mathcal{D}$ contains all cylinder sets, that it is closed under complements, and under countable disjoint unions. By the $\pi$-$\lambda$-Theorem A.1.1 $\mathcal{D}$ contains the $\sigma$-algebra generated by the cylinder sets, hence all Borel subset of $X^\infty$. This establishes uniqueness of the extension. \(\square\)

The proof (which is taken almost verbatim from [73, Chapter V.3]) makes critical use of the tightness property for finite measures on Polish spaces that says that we can approximate the measure of a Borel set arbitrarily well by compact sets, see Corollary A.3.4. It is also important that compact sets have the finite intersection property: if each finite intersection of a family of compact sets is nonempty, the intersection of the entire family cannot be empty. Thus the proof works in general Hausdorff spaces, provided the measures under consideration are tight.

The construction from above can be made use of when we work in Proposition A.3.7 in a compact scenario. We can get free us from that restrictive assumption using the Alexandrov embedding of Polish spaces into compact metric spaces that we also put to good use in section 4.2.2, when we transported a measure extension from a compact to a general Polish space. Following Parthasarathy [73] again, we will do that here as well.

**Proposition A.3.8** Let $X$ be a Polish space, $(\mu_n)_{n \in \mathbb{N}}$ be a projective system on $X$. Then there exists a unique projective limit $\mu_\infty \in \mathfrak{P}(X^\infty)$ for $(\mu_n)_{n \in \mathbb{N}}$. 

---

225
Proof. \(X\) is a dense measurable subset of a compact metric space \(\bar{X}\) by [51, Theorem 4.14]. Defining \(\bar{\mu}_n(B) := \mu_n(B \cap X^n)\) for the Borel set \(B \subseteq \bar{X}^n\) yields a projective system \((\bar{\mu}_n)_{n \in \mathbb{N}}\) on \(\bar{X}\) with a projective limit \(\bar{\mu}_\infty\) by Proposition A.3.7. Since by construction \(\bar{\mu}_\infty(X^\infty) = 1\), restrict \(\bar{\mu}_\infty\) to the Borel sets of \(X^\infty\), then the assertion follows. \(\dagger\)

Our interest in this construction comes from stochastic relations that may form a projective system. We will show now that there exists a stochastic relation which may be thought as the (pointwise) projective limit.

Corollary A.3.9 Let \(X\) be a Polish space, and assume that \(J^{(n)} : X \rightsquigarrow X^n\) is a stochastic relation for each \(n \in \mathbb{N}\) such that the sequence \((J^{(n)}(x))_{n \in \mathbb{N}}\) forms a projective system on \(X\) for each \(x \in X\). Then there exists a unique stochastic relation \(J_\infty : X \rightsquigarrow X^\infty\) such that \(J_\infty(x)\) is the projective limit of \((J^{(n)}(x))_{n \in \mathbb{N}}\) for each \(x \in X\).

Proof. Let for \(x\) fixed \(J_\infty(x)\) be the projective limit of the projective system \((J^{(n)}(x))_{n \in \mathbb{N}}\).

By Proposition 2.2.9 we need to show that the map \(x \mapsto J_\infty(x)(B)\) is measurable for every \(B \in \mathcal{B}(X^\infty)\).

1. In fact, consider

\[
\mathcal{D} := \{B \in \mathcal{B}(X^\infty) \mid x \mapsto J_\infty(x)(B) \text{ is measurable}\}
\]

then the general properties of measurable functions imply that \(\mathcal{D}\) is a \(\sigma\)-algebra on \(X^\infty\).

Take a cylinder set \(B = B_0 \times \prod_{j > k} X \) with \(B_0 \in \mathcal{B}(X^k)\) for some \(k \in \mathbb{N}\), then, by the properties of the projective limit, we have \(J_\infty(x)(B) = J^{(k)}(x)(B_0)\). But \(x \mapsto J^{(k)}(x)(B_0)\) constitutes a measurable function on \(X\). Consequently, \(B \in \mathcal{D}\), and so \(\mathcal{D}\) contains the cylinder sets which generate \(\mathcal{B}(X^\infty)\). Consequently, measurability is established for each Borel set \(B \subseteq X^\infty\). \(\dagger\)

Constructing a Projective System. Finally, we construct a projective system from a sequence \((J_n)_{n \in \mathbb{N}}\) of stochastic relations \(J_n : X \rightsquigarrow X\). Suppose that \(J_n\) given at step \(n\) the probability for a state transition, then \(J^{(n)}\) should model the state transitions for the \(n + 1\)-st step taking the history of steps \(1, \ldots, n\) into account (neglecting this history, we could model \(J^{(n)}\) just as the product of \(J_1, \ldots, J_n\)).

Well, then: put \(J^{(1)}(x) := J_1(x)\), and suppose that \(J^{(n)} : X \rightsquigarrow X^n\) has already been defined. Put for \(D \in \mathcal{B}(X^{n+1})\) and \(x \in X\)

\[
J^{(n+1)}(x)(D) := \int_{X^n} J_{n+1}(x_n)(D_{(x_1, \ldots, x_n)}) J^{(n)}(x)(d(x_1, \ldots, x_n)).
\]

This yields the probability that \(\langle x_1, \ldots, x_{n+1}\rangle \in D\) conditioned on starting in \(x \in X\): if at steps \(1\ldots n\) the system was in states \(\langle x_1, \ldots, x_n\rangle\), then \(J_{n+1}(x_n)(D_{(x_1, \ldots, x_n)})\) gives the probability for \(x_{n+1} \in D_{(x_1, \ldots, x_n)}\), or, equivalently, \(\langle x_1, \ldots, x_{n+1}\rangle \in D\). Because we know for a Borel set \(B \in \mathcal{B}(X^n)\) that

\[
(B \times X)_{(x_1, \ldots, x_n)} = \begin{cases} X, & (x_1, \ldots, x_n) \in B \\ \emptyset, & \text{otherwise,} \end{cases}
\]

and because \(J_{n+1}(x)(X) = 1\) holds for all \(x \in X\), we see that

\[
J^{(n+1)}(x)(B \times X) = J^{(n)}(x)(B)
\]

holds. Thus we have defined a projective system which we call the projective system associated with \((J_n)_{n \in \mathbb{N}}\).
A.3.4 Universal Measurability

Let $\mu \in \mathcal{S}(X, \mathcal{A})$ be a sub-probability on the measurable space $(X, \mathcal{A})$, then $A \subseteq X$ is called $\mu$-measurable iff there exist $M_1, M_2 \in \mathcal{A}$ with $M_1 \subseteq A \subseteq M_2$ and $\mu(M_1) = \mu(M_2)$. The $\mu$-measurable subsets of $X$ form a $\sigma$-algebra $\mathcal{M}_\mu(\mathcal{A})$. The $\sigma$-algebra $\mathcal{U}(\mathcal{A})$ of universally measurable sets is defined by

$$\mathcal{U}(\mathcal{A}) := \bigcap \{\mathcal{M}_\mu(\mathcal{A}) \mid \mu \in \mathcal{S}(X, \mathcal{A})\}$$

(in fact, one considers usually all finite or $\sigma$-finite measures; these definitions lead to the same universally measurable sets). If $f : X_1 \to X_2$ is an $\mathcal{A}_1$-$\mathcal{A}_2$-measurable map between the measurable spaces $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$, then it is well known that $f$ is also $\mathcal{U}(\mathcal{A}_1)$-$\mathcal{U}(\mathcal{A}_2)$-measurable; the converse does not hold, and one usually cannot conclude that a map $g : X_1 \to X_2$ which is $\mathcal{U}(\mathcal{A}_1)$-$\mathcal{A}_2$-measurable is also $\mathcal{A}_1$-$\mathcal{A}_2$-measurable.

**Lemma A.3.10** Let $X$ be a Polish, and $Y$ a second countable metric space. If $f : X \to Y$ is a surjective Borel map, so is $\mathcal{S}(f) : \mathcal{S}(X) \to \mathcal{S}(Y)$.

**Proof**

1. From [5, Theorem 3.4.3] we find a map $g : Y \to X$ such that $f \circ g = id_Y$ and $g$ is $\mathcal{U}(B(Y)) - \mathcal{U}(B(X))$-measurable.

2. Let $\nu \in \mathcal{S}(Y)$, and define $\mu := \mathcal{S}(g)(\nu)$, then $\mu \in \mathcal{S}(X, \mathcal{U}(B(X)))$ by construction. Restrict $\mu$ to the Borel sets on $X$, obtaining $\mu_0 \in \mathcal{S}(X, B(\mathcal{X}))$. Since we have for each set $B \subseteq Y$ the equality $g^{-1}[f^{-1}[B]] = B$, we see that for each $B \in B(Y)$

$$\mathcal{S}(f)(\mu_0)(B) = \mu_0(f^{-1}[B]) = \mu(f^{-1}[B]) = \nu(g^{-1}[f^{-1}[B]]) = \nu(B)$$

holds. $\dashv$
Appendix B

Notations etc.

Contents

<table>
<thead>
<tr>
<th>B.1 Categories</th>
<th>229</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.2 Spaces</td>
<td>230</td>
</tr>
<tr>
<td>B.3 Other</td>
<td>230</td>
</tr>
</tbody>
</table>

B.1 Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>All sets with maps</td>
<td>3</td>
</tr>
<tr>
<td>Meas</td>
<td>Measurable spaces with measurable maps</td>
<td>2.2</td>
</tr>
<tr>
<td>cPol</td>
<td>Polish spaces with continuous maps</td>
<td>3.1</td>
</tr>
<tr>
<td>BPol</td>
<td>Polish spaces with Borel maps</td>
<td>3.1</td>
</tr>
<tr>
<td>Anl</td>
<td>Analytic spaces with Borel maps</td>
<td>3.1</td>
</tr>
<tr>
<td>Stoch</td>
<td>Stochastic relations over measurable spaces</td>
<td>3.1</td>
</tr>
<tr>
<td>PolStoch</td>
<td>Stochastic relations over Polish spaces</td>
<td>3.1</td>
</tr>
<tr>
<td>anStoch</td>
<td>Stochastic relations over analytic spaces</td>
<td>3.1</td>
</tr>
<tr>
<td>StrConv</td>
<td>Positive convex structures with continuous affine maps</td>
<td>3.1</td>
</tr>
<tr>
<td>GPart</td>
<td>G-partitions with partition respecting continuous maps</td>
<td>3.1</td>
</tr>
<tr>
<td>Alg</td>
<td>Algebras for the Giry monad with algebra morphisms</td>
<td>3.1</td>
</tr>
<tr>
<td>pAlg</td>
<td>Subcategory of Alg for probabilistic objects</td>
<td>3.1</td>
</tr>
<tr>
<td>GTrip</td>
<td>G-triplets with G-triplet morphisms</td>
<td>3.1</td>
</tr>
<tr>
<td>C(a,b)</td>
<td>Morphisms $a \to b$ in category C</td>
<td>3.6</td>
</tr>
<tr>
<td>Prob</td>
<td>Measurable spaces with probability measures</td>
<td>4.1</td>
</tr>
<tr>
<td>PolProb</td>
<td>Full subcategory of Prob based on Polish spaces</td>
<td>4.1</td>
</tr>
<tr>
<td>pKripke</td>
<td>Stochastic Kripke models</td>
<td>6.1.3</td>
</tr>
</tbody>
</table>
B.2 Spaces

\[ \mathcal{F}(\mathbb{N}, \mathbb{N}) \text{ \text{-} } \mathcal{N}\mathcal{B}(\mathbb{R})\text{-measurable and bounded functions } f : \mathbb{N} \to \mathbb{R} \]
\[ \bigotimes_{i \in I}(X_i, A_i) \text{ Product of the measurable spaces } (X_i, A_i)_{i \in I} \]
\[ \bigoplus_{i \in I}(X_i, A_i) \text{ Sum of the measurable spaces } (X_i, A_i)_{i \in I} \]
\[ \mathcal{S}(\mathbb{N}, \mathbb{N}) \text{ Sub-probability measures on the measurable space } (\mathbb{N}, \mathbb{N}) \]
\[ \mathfrak{P}(\mathbb{N}, \mathbb{N}) \text{ Probability measures on the measurable space } (\mathbb{N}, \mathbb{N}) \]
\[ \mathcal{C}(X) \text{ All bounded continuous functions } X \to \mathbb{R} \]
\[ X^\infty \text{ Infinite product of } X \text{ with itself} \]

B.3 Other

\[ \cdot \text{ } \text{ Natural transformation between functors} \]
\[ \text{cl} \text{ } \text{ Topological closure} \]
\[ \sim \text{ } \text{ Kleisli morphism} \]
\[ \varepsilon, m \text{ } \text{ Unit and multiplication of a monad} \]
\[ \otimes_H \text{ } \text{ H-product} \]
\[ \Lambda(\mathcal{G}) \text{ } \text{ Number of strata in a dag} \]
\[ i(\cdot, n) \text{ } \text{ Input to node } n \]
\[ o(\cdot, n) \text{ } \text{ Output from node } n \]
\[ a(\cdot, n) \text{ } \text{ Work being done in node } n \]
\[ g(\cdot, S) \text{ } \text{ Flow into stratum } S \]
\[ A(\cdot, S) \text{ } \text{ Work being done in stratum } S \]
\[ P(\mathcal{G}) \text{ } \text{ Work being done in pipeline } \mathcal{G} \]
\[ [\mathcal{G}_1, \gamma] +_\tau [\mathcal{G}_2, \chi] \text{ } \text{ } \tau\text{-concatenation of PF-systems} \]
\[ \mathcal{G}_1|\mathcal{G}_2 \psi \text{ } \psi\text{-replacement of node } n \text{ through } \mathcal{G}_2 \]
\[ K^\bullet : \mathcal{G}(X) \to \mathcal{G}(Y) \text{ } \text{ Morphism associated with } K : X \sim \sim Y \]
\[ \mathcal{C}[\cdot] \text{ } \text{ Semantic function} \]
\[ [\cdot]_\rho \text{ } \text{ Equivalence class for equivalence relation } \rho \]
\[ \eta_\rho \text{ } \text{ Factor map for equivalence relation } \rho \]
\[ X/\rho \text{ } \text{ Factor space for equivalence relation } \rho \]
\[ T/\rho \text{ } \text{ Factor topology for equivalence relation } \rho \]
\[ \ker(f) \text{ } \text{ Kernel of map } f \]
\[ \delta_{i,j} \text{ } \text{ Kronecker's } \delta \]
\[ \Omega \text{ } \text{ Set of positive convex coefficients} \]
\[ \Omega_c \text{ } \text{ Set of convex coefficients} \]
\[ x \bullet x' \text{ } \text{ Inner product of two vectors} \]
\[ \mathcal{C}_Q[\cdot] \text{ } \text{ Derandomization semantics} \]
\[ \ell(R) \text{ } \text{ Equivalence relation generated from } R \]
\[ [g_1\|g_2] \text{ } \text{ Common product image} \]
\[ (X, \alpha) \text{ } \text{ Trace of the } \alpha\text{-invariant Borel sets on } \alpha \]
\[ (\times_{n \in \mathbb{N}} \rho_n) \text{ } \text{ Infinite product of equivalence relations} \]
\[ \mathcal{I}N\mathcal{V}(\mathcal{B}(X), \rho) \text{ } \text{ } \rho\text{-invariant Borel sets of an analytic space } X \]
\( \rho + \sigma \) Sum of the equivalence relations \( \rho \) and \( \sigma \)  
\( \tau \cdot \rho \) Collects smooth equivalence relations \( \tau \) and \( \rho \)  
\( \Delta_{X \times U_X} \) Identity relation resp. universal relation on \( X \)  
\( \ker(f) \) Kernel of morphism \( f \)  
\( K_{\alpha, \beta}, K/c \) Factor relation  
\( c \cdot d \) Collects congruences \( c \) and \( d \)  
\( (\alpha, \beta) \preceq (\alpha', \beta') \) Refinement of congruences  
\( c \propto c' \) Proportionality of congruences  
\( K \oplus K' \) Direct sum of stochastic relations \( K \) and \( K' \)  
\( \nu << \mu, M << \mu \) Absolute continuity w.r.t. a measure  
\( V_n \) All permutations on \{1, \ldots, n\}  
\( H_n \) All heaps in \( V_n \)  
\( \text{Mod}_b(\tau, P) \) Basic modal language  
\( \text{Mod}_s(\tau, P) \) Extended modal language  
\( \models \) Satisfaction relation  
\( \rho(\Delta) \) Arity of modal operator \( \Delta \)  
\( Th_\mathcal{R}(s) \) Theory for state \( s \) with Kripke model \( \mathcal{R} \)  
\( K \circ \mathcal{R} \) \( K \) refines \( \mathcal{R} \)  
\( \text{supp}(\mu) \) Support of probability measure \( \mu \)  
\( K \sim K' \) Equivalence of Kripke models \( K \) and \( K' \)  
\( \mathcal{L}_P \) All state formulas in \( \text{CSL} \)  
\( S_\phi(\varphi) \) Steady state operator in \( \text{CSL} \)  
\( \mathcal{P} \) Path quantifier in \( \text{CSL} \)  
\( \mathcal{X}^I \varphi \) Next operator in \( \text{CSL} \)  
\( \varphi_1 U^I \varphi_2 \) Until operator in \( \text{CSL} \)  
\( \sigma @ t \) State occupied by \( \sigma \) at time \( t \)  
\( \text{PATHS} \) All infinite paths in \( \text{CSL} \)  
\( \text{wrap}(F) \) Closure of a set \( F \) of formulas  
\( \text{ext}(F) \) Extension of a set \( F \) of formulas  
\( \sigma(M_0) \) \( \sigma \)-algebra generated by \( M_0 \)  
\( \chi_A \) Indicator function for set \( A \)  
\( B(X) \) Borel sets of \( X \), \( X \) is a topological or an analytic space  
\( \exists F(C), \forall F(C) \) Weak and strong inverse of set-valued map \( F \)  
\( A^* \) Weak*-\( \sigma \)-algebra on \( \mathcal{G}(X, A) \)  
\( d_P \) Prohorov metric on \( \mathcal{G}(X) \) for the metric space \( X \)  
\( \delta_a \) Dirac measure on the point \( a \)  
\( \rightarrow_w \) Weak convergence for probability measures  
\( D_x, D_y \) Horizontal and vertical cuts of \( D \subseteq X \times Y \)  
\( U(A) \) \( \sigma \)-algebra of universally measurable sets over \( \sigma \)-algebra \( A \)
Bibliography


233


Bibliography


### Index

**$G_\delta$-set** ........................................ 218

**$\sigma$-algebra** ........................................ 118

- **Borel set** ........................................ 216
  - invariant ........................................ 118
  - common events .................................... 122, 138
  - final ........................................ 216
  - initial ........................................ 216
  - product ........................................ 216
  - sum ........................................ 216
  - trace ........................................ 219

**Abowd, G.** ........................................ 53, 68

**Abramsky, S.** .......................................... 159, 161, 170

**absolutely continuous** ................................ 161

**algorithm** ........................................ 170

- **Floyd’s** ........................................ 170
- randomized ........................................ 94
- Williams ........................................ 167

**algorithms** ........................................ 165

- **average behavior** .................................. 165
- **Allen, R.** .......................................... 53, 68
- Arbab, F. ........................................ 11, 68
- Arbib, M. A. ........................................ 68, 92, 93
- Arveson, W. ........................................ 10, 119, 170, 219, 227
- **atom** ........................................ 129, 139, 220
- Aumann, G. .......................................... 148
- axiom of choice ..................................... 99

**2-bisimulation** .......................................... 11, 145

**smooth** ........................................ 145

**weak** ........................................ 145

**mediator** ........................................ 138

**Blackburn, P.** ........................................ 5, 11, 99, 175, 177

**Blute, R.** ........................................ 159, 161, 170

**Bowman, H.** ........................................ 54, 212

**Broy, M.** ........................................ 49, 54, 68

**Bruni, R.** ........................................ 68

**Bryans, J.** ........................................ 54, 212

**Carboni, A.** ........................................ 54, 68

**category** ........................................ 27

**comma** ........................................ 117

**extensive** ........................................ 117

**lextensive** .......................................... 117

**Change of Variable formula** ................................ 19

**Cirstea, C.** .......................................... 171

**Clarke, E. M.** ........................................ 212

**coalgebra** ........................................ 5, 26

**coinduction** ........................................ 11, 165, 171

**congruence** ........................................ 10, 132

**equivalence** ........................................ 139

**non-trivial** ........................................ 133

**plain** ........................................ 149

**proportionality** ..................................... 139

**refinement** .......................................... 136

**D’Argenio, P.R.** ...................................... 54

**Baier, C.** ........................................ 192, 193, 212

**Barbosa, L. M.** ...................................... 36, 54, 68

**Barr, M.** ........................................ 19

**barycenter** ........................................ 88

**Billingsley, P.** ...................................... 24, 59, 215, 222

**bisimilar** ........................................ 137

**bisimulation** ........................................ 10, 137

**connector** ........................................ 53

**cut** ........................................ 223

**horizontal** .......................................... 223

**vertical** ........................................ 223

**D’Argenio, P.R.** ...................................... 54
<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>de Rijke, M.</td>
<td>5, 11, 99, 175, 177</td>
<td></td>
</tr>
<tr>
<td>decision</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>randomized</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>DeLine, R.</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Giry, M.</td>
<td>19, 22, 67</td>
<td></td>
</tr>
<tr>
<td>G-partition</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>G-triplet</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Garlan, D.</td>
<td>28, 53, 68</td>
<td></td>
</tr>
<tr>
<td>Gribskov, M.</td>
<td>11, 55, 69, 109, 113</td>
<td></td>
</tr>
<tr>
<td>Gumm, H. P.</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>de Rijke, M.</td>
<td>5, 11, 99, 175, 177</td>
<td></td>
</tr>
<tr>
<td>decision</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>randomized</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>DeLine, R.</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Giry, M.</td>
<td>19, 22, 67</td>
<td></td>
</tr>
<tr>
<td>G-partition</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>G-triplet</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Garlan, D.</td>
<td>28, 53, 68</td>
<td></td>
</tr>
<tr>
<td>Gribskov, M.</td>
<td>11, 55, 69, 109, 113</td>
<td></td>
</tr>
<tr>
<td>Gumm, H. P.</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Fiadeiro, J. L.</td>
<td>29, 49, 53, 68</td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>54, 68</td>
<td></td>
</tr>
<tr>
<td>Eilenberg-Moore algebra</td>
<td>9, 72</td>
<td></td>
</tr>
<tr>
<td>morphism</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Elstrodt, J.</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Engelking, R.</td>
<td>8, 86</td>
<td></td>
</tr>
<tr>
<td>Engels, G.</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>equivalence relation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>countably generated</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>factor space</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>positive convexity</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>smooth</td>
<td>79, 117</td>
<td></td>
</tr>
<tr>
<td>spawning</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>Fedorchuk, V. V.</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Fiadeiro, J. L.</td>
<td>29, 49, 53, 68</td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>54, 68</td>
<td></td>
</tr>
<tr>
<td>Fremlin, D. H.</td>
<td>88, 89, 215</td>
<td></td>
</tr>
<tr>
<td>Freyd, P.</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Fuchssteiner, B.</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>indicator</td>
<td>216</td>
<td></td>
</tr>
<tr>
<td>step</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>functor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hom-set</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>monoidal</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>power set</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>probability</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Gjergji, V.</td>
<td>28, 53, 68</td>
<td></td>
</tr>
<tr>
<td>Graham, R. E.</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>graph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q-extension</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>directed acyclic</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>stratified</td>
<td>30, 36</td>
<td></td>
</tr>
<tr>
<td>Grumberg, O.</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>Gumm, H. P.</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Gupta, V.</td>
<td>67, 170</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>5, 215</td>
<td></td>
</tr>
<tr>
<td>Halmos, P. R.</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>Hausmann, J. H.</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Haverkort, B.</td>
<td>192, 193, 212</td>
<td></td>
</tr>
<tr>
<td>heap</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>Heckmann, R.</td>
<td>68, 91</td>
<td></td>
</tr>
<tr>
<td>Hennessy, M.</td>
<td>11, 170, 178, 212</td>
<td></td>
</tr>
<tr>
<td>Hermanns, H.</td>
<td>192, 193, 212</td>
<td></td>
</tr>
<tr>
<td>Hewitt, E.</td>
<td>20, 215</td>
<td></td>
</tr>
<tr>
<td>Himmelberg, C. J.</td>
<td>220, 221</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>5, 215</td>
<td></td>
</tr>
<tr>
<td>Halmos, P. R.</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>Hausmann, J. H.</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Haverkort, B.</td>
<td>192, 193, 212</td>
<td></td>
</tr>
<tr>
<td>heap</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>Heckmann, R.</td>
<td>68, 91</td>
<td></td>
</tr>
<tr>
<td>Hennessy, M.</td>
<td>11, 170, 178, 212</td>
<td></td>
</tr>
<tr>
<td>Hermanns, H.</td>
<td>192, 193, 212</td>
<td></td>
</tr>
<tr>
<td>Hewitt, E.</td>
<td>20, 215</td>
<td></td>
</tr>
<tr>
<td>Himmelberg, C. J.</td>
<td>220, 221</td>
<td></td>
</tr>
<tr>
<td>Jacob, J.</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Jagadeesan, R.</td>
<td>67, 170</td>
<td></td>
</tr>
<tr>
<td>Jerschov, M.</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Jones, C.</td>
<td>60, 68</td>
<td></td>
</tr>
<tr>
<td>Joyal, A.</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>jungle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>measure-theoretic</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Katoen, J-P.</td>
<td>192, 193, 212</td>
<td></td>
</tr>
<tr>
<td>Keisler, H. J.</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>kernel</td>
<td>78, 132</td>
<td></td>
</tr>
<tr>
<td>Klein, D. V.</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Kleisli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>category</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>morphism</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Kleisli construction</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Knuth, D. E.</td>
<td>166–168</td>
<td></td>
</tr>
<tr>
<td>Kozen, D. E.</td>
<td>55, 68</td>
<td></td>
</tr>
<tr>
<td>Kripke model</td>
<td>5, 215</td>
<td></td>
</tr>
<tr>
<td>degenerate</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>equivalent</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>image-finite</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>morphism</td>
<td>183</td>
<td></td>
</tr>
</tbody>
</table>
Index

nondeterministic .................................. 175
refinement ........................................ 180
stochastic ........................................ 176
strong morphism ................................ 185
strongly bisimilar ................................ 185
Kuratowski, K ................................... 218

L
labelled transition system ......................... 178
Lack, S ............................................ 54, 68
Lajios, G .......................................... 27, 54, 68, 112
Larsen, K. G ...................................... 11, 67, 177, 212
Lang, S ............................................ 136, 149, 171
language
  modal
    basic ......................................... 175
    basic temporal ................................ 178
    extended ..................................... 175
    negation free ................................ 175
Larsen, K. G ...................................... 11, 67, 177, 212
Lindstrøm, T ..................................... 7, 26, 181
Loève, M .......................................... 24
logic
  CSL ............................................. 191
  $F$-bisimulation ................................. 209
  closure ........................................ 202
  DP-condition .................................. 203
  extension ..................................... 211
  rate model ................................... 192, 197
  Zeno paths .................................... 197
  pCTL* .......................................... 190
continuous stochastic ................................ 11
modal ............................................. 5
arrow ............................................ 179
Hennessy-Milner ................................ 178
negation free .................................... 180
similarity type ................................ 175
Lohmann, M ....................................... 68
Lusky, W ......................................... 89

M
MacLane, S ........................................ 9, 19, 27, 32, 35, 36, 67, 71, 73, 90
MacQueen, D ................................... 55, 61
Maibaum, T ...................................... 29, 53, 68
Manes, E. G ..................................... 68, 92, 93
map
measurable ....................................... 216
marginal distribution ............................. 158
Markov transition system .......................... 11, 67, 178
Markov transition systems ......................... 170
McIver, A ..................................... 26, 91
measurable
  jointly .......................................... 223
  monoid ......................................... 21
  rectangle ...................................... 216
  universal ...................................... 112, 227
measure
  probability .................................... 221
  product ........................................ 223
  projective
    limit .......................................... 225
    system ........................................ 224
    sub-probability ................................ 221
    tight ......................................... 222
Medvidovics, N .................................. 68
Milner, R ......................................... 11, 170, 178, 212
Mitchell, J. C .................................. 65
model
  qualitative ..................................... 174
  quantitative ................................... 174
Moggi, E ......................................... 27, 29, 36, 60
monad ............................................. 17
  $\#$-condition ................................ 42
  Giry ........................................... 9
Manes ............................................. 9, 18
mediating transformation .......................... 33
multiplication .................................... 17
product compatibility ................................ 33
strong ........................................... 36
tensorial strength ................................ 36
unit .............................................. 17
Monniaux, D ..................................... 69
Montanari, U .................................... 68
Morgan, C ........................................ 26, 91
Moss, L. S ....................................... 171

N
Nielsen, M ....................................... 170
number
  harmonic ....................................... 166
  Stirling ....................................... 166
### Index

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Object analytic</td>
</tr>
<tr>
<td>142</td>
<td>Block</td>
</tr>
<tr>
<td>154</td>
<td>Final</td>
</tr>
<tr>
<td>154</td>
<td>Simple</td>
</tr>
<tr>
<td>132</td>
<td>Factor</td>
</tr>
<tr>
<td>27</td>
<td>Polski</td>
</tr>
<tr>
<td>189</td>
<td>Polish</td>
</tr>
<tr>
<td>166</td>
<td>Patashnik</td>
</tr>
<tr>
<td>212</td>
<td>Peled D. A.</td>
</tr>
<tr>
<td>81</td>
<td>P-G triplets</td>
</tr>
<tr>
<td>69</td>
<td>Pierro A. Di</td>
</tr>
<tr>
<td>27</td>
<td>Pipeline filter</td>
</tr>
<tr>
<td>27</td>
<td>Pipeline pipe</td>
</tr>
<tr>
<td>27</td>
<td>UNIX pipes</td>
</tr>
<tr>
<td>37</td>
<td>Pipeline system</td>
</tr>
<tr>
<td>68</td>
<td>Pleumann J.</td>
</tr>
<tr>
<td>68</td>
<td>Plotkin G.</td>
</tr>
<tr>
<td>82</td>
<td>Positive convexity</td>
</tr>
<tr>
<td>82</td>
<td>Affine map</td>
</tr>
<tr>
<td>75</td>
<td>Morphism</td>
</tr>
<tr>
<td>68</td>
<td>Tuples</td>
</tr>
<tr>
<td>84</td>
<td>Probabilistic powerdomain</td>
</tr>
<tr>
<td>12</td>
<td>Projective limit</td>
</tr>
<tr>
<td>100</td>
<td>Pullback lifting</td>
</tr>
<tr>
<td>10</td>
<td>Semi-</td>
</tr>
<tr>
<td>111</td>
<td>Weak</td>
</tr>
<tr>
<td>27</td>
<td>Pumplun D.</td>
</tr>
<tr>
<td>68</td>
<td>Redmiles D. F.</td>
</tr>
<tr>
<td>68</td>
<td>Regular conditional distribution</td>
</tr>
<tr>
<td>107</td>
<td>Distribution</td>
</tr>
<tr>
<td>157</td>
<td>Probability</td>
</tr>
<tr>
<td>106</td>
<td>Relation C-measurable</td>
</tr>
<tr>
<td>181</td>
<td>Measurable</td>
</tr>
<tr>
<td>221</td>
<td>Castaing representation</td>
</tr>
<tr>
<td>19</td>
<td>Nondeterministic</td>
</tr>
<tr>
<td>174</td>
<td>Refinement</td>
</tr>
<tr>
<td>176</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>19</td>
<td>Set-theoretic</td>
</tr>
<tr>
<td>24</td>
<td>Stochastic</td>
</tr>
<tr>
<td>111</td>
<td>Weak inverse</td>
</tr>
<tr>
<td>220</td>
<td>Weakly measurable</td>
</tr>
<tr>
<td>53</td>
<td>Reo</td>
</tr>
<tr>
<td>68</td>
<td>Robbins J. E.</td>
</tr>
<tr>
<td>68</td>
<td>Rosenblum D. S.</td>
</tr>
<tr>
<td>53</td>
<td>Ross T. L.</td>
</tr>
<tr>
<td>53</td>
<td>Rutten J. M. M.</td>
</tr>
<tr>
<td>67</td>
<td>Schal M.</td>
</tr>
<tr>
<td>28</td>
<td>Schmidt F.</td>
</tr>
<tr>
<td>68</td>
<td>Schröder J.</td>
</tr>
<tr>
<td>171</td>
<td>Schröder T.</td>
</tr>
<tr>
<td>26</td>
<td>Seidel K.</td>
</tr>
<tr>
<td>220</td>
<td>Selector</td>
</tr>
<tr>
<td>98</td>
<td>Measurable</td>
</tr>
<tr>
<td>98</td>
<td>Semadeni Z.</td>
</tr>
<tr>
<td>218</td>
<td>Clopren</td>
</tr>
<tr>
<td>218</td>
<td>Co-analytic</td>
</tr>
<tr>
<td>220</td>
<td>Cylinder</td>
</tr>
<tr>
<td>125</td>
<td>Sethi R.</td>
</tr>
<tr>
<td>55</td>
<td>Shalit S.</td>
</tr>
<tr>
<td>91</td>
<td>Shaw M.</td>
</tr>
<tr>
<td>28</td>
<td>Shiryaev A. N.</td>
</tr>
<tr>
<td>28</td>
<td>Software architecture</td>
</tr>
<tr>
<td>49</td>
<td>t-concatenation</td>
</tr>
<tr>
<td>28</td>
<td>Components</td>
</tr>
<tr>
<td>28</td>
<td>Connectors</td>
</tr>
</tbody>
</table>

243
filter ........................................ 29
glass-box refinement .................. 49
pipeline .................................. 29
stratified .................................. 38
style ...................................... 28
system evolution .......................... 48
space
analytic .................................. 8, 218
  Borel sets ................................ 219
connected ................................ 86
measurable ................................ 215
separable ................................ 219
Polish .................................... 8, 217
  Borel sets ................................ 215, 217
  Standard Borel ......................... 220
Spivey, J. M. .............................. 53
Stølen, K. .................................. 49, 54, 68
Stromberg, K. R. .......................... 20, 215
support of a probability ............... 180
Swirszcz, T. ................................ 94
system
  reactive .................................. 12
T
Tarski, A. .................................. 161
Taylor, P. .................................. 54
testing ..................................... 67, 202
Theorem
 $\pi$-$\lambda$ .................................. 216
Alexandrov ......................... 104, 218, 225
Blackwell-Mackey ............. 106, 120, 220
Bounded Convergence .................. 134
Fubini .................................. 194
Fubinito .................................. 159
Hahn-Banach ......................... 103
Hennessy-Milner ..................... 170, 187
Kleene-Knaster-Tarski ............. 207
Monotone Convergence ............. 221
Portmanteau ................................ 222
Riesz Representation ................. 103
Second Isomorphism .................. 136
Souslin ................................... 120, 121, 220
transformation ............................ 176

psi-replacement .............................. 50

U
UNITY ....................................... 68
upper-semicontinuity .................. 75

V
van Breugel, F. .................. 91
Veltmann, C. ...................... 28
Venema, Y. .......................... 5, 11, 99, 175, 177
Viglizzo, I. D. ......................... 171

W
Wagner, D. H. ..................... 221
Wagon, S. .................................. 7
Walters, R. .............................. 54, 68
weak
  $\ast$-$\sigma$-algebra ................. 222
topology ................................ 222
  Prohorov metric ...................... 222
Wells, C. .................................. 19
Wermelinger, M. ..................... 29, 49, 53, 68
Wiklicky, H. ......................... 69
Williams, J. W. J. .................. 168
Winkelstein, L. .................... 55
Worrell, J. .............................. 91

Y
Young, D. M. ......................... 53

Z
Zelesnik, G. ....................... 53
Zorn's Lemma ......................... 99