Markov Processes in Isabelle/HOL

Johannes Hölzl
Technical University of Munich, Germany
hoelzl@in.tum.de

Abstract

Markov processes with discrete time and arbitrary state spaces are important models in probability theory. They model the infinite steps of non-terminating programs with (not just discrete) probabilistic choice and form the basis for further probabilistic models. Their transition behavior is described by Markov kernels, i.e. measurable functions from a state to a distribution of states. Markov kernels can be composed in a monadic way from distributions (normal, exponential, Bernoulli, etc.), other Markov kernels, and even other Markov processes.

In this paper we construct discrete-time Markov processes with arbitrary state spaces, given the transition probabilities as a Markov kernel. We show that the Markov processes form again Markov kernels. This allows us to prove a bisimulation argument between two Markov processes and derive the strong Markov property. We use the existing probability theory in Isabelle/HOL and extend its capability to work with Markov kernels.

As application we construct continuous-time Markov chains (CTMCs). These are constructed as jump & hold processes, which are discrete-time Markov processes where the state space is a product of continuous holding times and discrete states. We prove the Markov property of CTMCs using the bisimulation argument for discrete-time Markov processes, and that the transition probability is the solution of a differential equation.

Categories and Subject Descriptors G.3 [*Probability and Statistics*]: Markov Processes

Keywords Isabelle/HOL, Interactive Theorem Proving, Probability Theory, Markov Kernels, Markov Processes, Continuous-Time Markov Chains

CPP '17, January 16 - 17, 2017, Paris, France

ACM 978-1-4503-4705-1/17/01…$15.00
DOI: http://dx.doi.org/10.1145/3018610.3018628

1. Introduction

Modelling stochastic processes often requires complicated constructions of probability measures. The Giry monad, a monad on probability measures, gives us a nice way to compose and transform probability measures. Discrete-time Markov processes are a further construction principle, which allow us to model infinite recursive behavior. In this paper we present a formalization which provides the components to model more complex stochastic processes. As stochastic processes we will present discrete-time Markov processes and continuous-time Markov chains.

A continuous-time Markov process on a discrete state space is called a *continuous-time Markov chain* (A *Markov chain* is a Markov process on a discrete state space). They are applied as models in queuing and reliability theory (Trivedi 2002) as well as resource management (Norris 1997), they can be used to represent stochastic process algebras (Clark et al. 2007), or model epidemics or genetic variations in biology (Norris 1997). For example in queuing theory, CTMCs describe the queuing behavior in a server-client architecture. As state we take the length of the request queue from the client to the server, and with a rate r_c a client submits a new request and with a rate r_s the server responds to a request.

This formalization provides us also the possibility to verify CTMC model checking algorithms. For this we formalize a core theorem for bounded reachability analysis on CTMCs: the backward equation (Baier et al. 2003).

Discrete-time Markov processes Discrete-time Markov processes are a generalization of discrete-time Markov chains which are models for finite probabilistic automata. We use a programming oriented way to describe discrete-time Markov processes: we model their behavior as recursive probabilistic programs. A process proc implements the following scheme:

$$\begin{aligned}&\mathsf{proc}\ x = \mathbf{do}\ \{\\ &\quad y \leftarrow K\ x\\ &\quad \omega \leftarrow \mathsf{proc}\ y\\ &\quad \mathsf{return}\ (y{\cdot}\omega)\\ &\}\end{aligned} \tag{1}$$

This is interpreted as follows: for an input parameter x, we first select a y from the distribution $K\ x$, then we recursively

generate a trace ω using proc y, and finally we return $y{\cdot}\omega$ the trace constructed by prepending y to ω. The trace ω is an infinite sequence of states.

Let's take a look at a more concrete example: a bounded random walk of a slowing particle.

$$\begin{aligned}&\mathsf{proc}\ (x,\sigma) = \textbf{do}\ \{\\ &\quad x' \leftarrow \mathsf{Normal}\ (x,\sigma)\\ &\quad x' := (\textbf{if}\ |x'| > 10\ \textbf{then}\ 0\ \textbf{else}\ x')\\ &\quad \omega \leftarrow \mathsf{proc}\ (x', \sigma/2)\\ &\quad \mathsf{return}\ ((x',\sigma/2){\cdot}\omega)\\ &\}\end{aligned}$$

The state of the Markov process contains the current position x and the variance for the next position σ. This information will also be stored in the trace. It is important to note that the distribution of x' is a continuous distribution: it is the normal distribution with mean x and variance σ. From this follows that the state space of all (x, σ) itself is continuous. In order to measure sets on continuous distributions we need to enforce that the set of x' we want to measure is measurable.

Markov kernels Therefore we must state everything in terms of measurable spaces and measurable functions, e.g. in Eq. (1): the functional K, the monadic operations, and also the process proc itself. Luckily, all these fit very nicely in the framework of measure theory and measurability. The measurable spaces and measurable functions between them form a category, i.e. they are closed under composition and the identity function is measurable. There are measurable spaces on all involved spaces: on the real numbers, on infinite sequences, on countable sets, and also on the set of probability measures on a fixed measurable space.

The last measurable space is important for *Markov kernels* (also called *stochastic relations* (Doberkat 2007)): measurable function from states into probability measures on states. For each Markov kernel we construct a discrete-time Markov process proc, fulfilling Eq. (1).

A concept related to Markov kernels is *conditional probability*: the *conditional probability distribution* $\Pr(\cdot|X = x)$ is a Markov kernel in the variable x. Moreover the defining property of conditional probability distributions is nicely expressed using the Markov kernel representation. A Makrov kernel K is a conditional probability of a probability space $\mathbb{P}$ for a random variable X (i.e. $K\ x = \Pr(\cdot|X = x)$) if

$$\mathbb{P} = \textbf{do}\ \{\omega \leftarrow \mathbb{P};\ \omega' \leftarrow K\ (X\ \omega);\ \mathsf{return}\ \omega'\}. \qquad (2)$$

We introduce Markov processes $M\ s$ (where s is the starting state) which are also conditional probability distributions:

$$M\ s = \textbf{do}\ \{\omega \leftarrow M\ s;\ \omega' \leftarrow M\ \omega_t;\ \mathsf{return}\ (\omega \cdot_t \omega')\}. \qquad (3)$$

This relates to Eq. (2), where the probability space $\mathbb{P} = M\ s$ and the random variable $X\ \omega = \omega_t$. We write ω_t for the state at time t and $\omega \cdot_t \omega'$ for the concatenation of the first t states of ω with ω'.

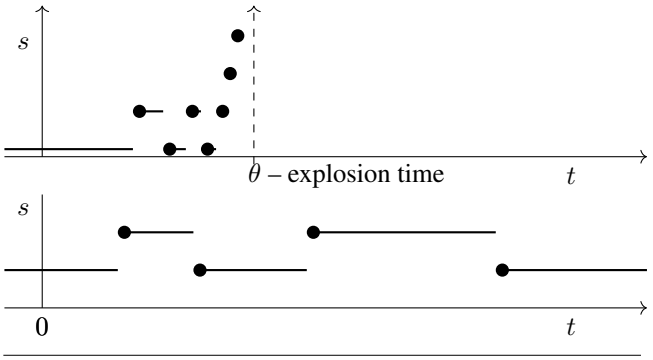

Figure 1. Two traces of continuous-time Markov chains

Now, how do we present the Markov property in terms of Markov kernels? The Markov property for Markov processes is usually stated as:

$$\Pr(A \mid \forall t' \le t.\ X_{t'} = x_{t'}) = \Pr(A \mid X_t = x_t). \qquad (4)$$

where $0 \le t$, X is the sequence of random variables describing the Markov process, and A observes only $X_{t'}$ for $t' \ge t$, i.e. A is a formula with statements of the form $X_{t'} \in B$ where $t' \ge t$. Also $\Pr(\forall t' \le t.\ X_{t'} = x_{t'}) > 0$, otherwise the conditional probabilities are not well-defined. We avoid the latter condition, by multiplying both sides with that probability:

$$\begin{aligned}&\Pr(A \wedge \forall t' \le t.\ X_{t'} = x_{t'}) =\\ &\quad \Pr(A \mid X_t = x_t) \cdot \Pr(\forall t' \le t.\ X_{t'} = x_{t'}).\end{aligned}$$

But with a some massaging, this is a specialized form of Eq. (3). Remember that A only observes X_t for $t \ge t$, hence $\omega \cdot_t \omega' \in A$ is independent of ω. Hence Eq. (4) is a specialized form of Eq. (3), and in our formalization we proof the Markov properties only in form of Eq. (3).

Continuous-time Markov chains As an application of discrete-time Markov processes we formalize continuous-time Markov chains (CTMCs). CTMCs are automata where the transition from a state x to a state y is annotated with a *rate* $R\ x\ y$. The rate describes how often a transition is taken per time unit.

We are not only interested in the transition system, but also in the trace space, i.e. a probability measure on traces $\mathbb{R} \to S$ mapping time to states. For example in the queuing system we want to know if the queue length converges when the time goes to infinity. Such traces are described by jump & hold processes which probabilistically choose a state and a sojourn time, see Figure 1. The bullets at the beginning of a jump mean that the trace is in the new state at the time of the jump. While the lower trace is going to infinity (at least in the visible part), the upper trace is exploding: its jump times converge towards a finite value θ.

An alternative model for CTMCs would allow parallel choices for each y with $R\ x\ y > 0$. Both models are equal: going from x to y with probability $R\ x\ y / \sum_y R\ x\ y$ is the

same as starting all transitions in parallel and then choosing the state with the shortest sojourn time.

The transition function $p_{x,y}(t) = \Pr_x(X(t) = y)$ describes the probability for a Markov chain started in x to be in y at time t. The transition function p is characterized by the *backward equation*, a differential equation depending on R. The backward equation plays a central role in analyzing CTMCs, they are at the core of model checking CTMCs (Baier et al. 2003).

Overview In the next section relate our formalization with existing work in Isabelle/HOL and other theorem provers. In Section 3 we will give a short introduction into the existing measure and probability theory in Isabelle/HOL. In Section 4 we show how we construct Markov processes, derive their iteration and uniqueness rules, and finally prove their strong Markov property. For the strong Markov property we also introduce stopping times, a special kind of random variables. In Section 5 we introduce continuous-time Markov chains, give an alternative model for their construction, prove their Markov property, and introduce the transition function p. We verify that the transition function is a so-called matrix semi-group and show its backward equation. Finally, we discuss the formalization in Section 6.

We use the same Isabelle theories as the formalization of discrete-time Markov chains in (Hölzl 2016). (Hölzl 2016) is restricted to discrete state spaces, while this paper presents a general development for arbitrary measurable state spaces. Both formalizations are in the Markov Models entry in the AFP (Hölzl and Nipkow 2012a).

2. Related Work

The work presented in this paper is founded on Isabelle's measure and probability theory presented in (Hölzl 2013; Hölzl and Nipkow 2012b,c). The previous work constructs finite-state Markov chains to verify probabilistic model checking. This was extended in (Hölzl 2016) to countable (possibly infinite) discrete state spaces, and further analysis concepts such as (positive) recurrence, the period of a state and stationary distributions. While the discrete-time Markov processes presented in this paper are more general we did not yet port the existing theories in the AFP (Hölzl and Nipkow 2012a) to our new formalization. The formalizations presented in the preliminaries in Section 3 were already introduced for (Eberl et al. 2015) and (Hölzl 2016). Since the work in (Hölzl 2016) focuses on Makrov chains (i.e. Makrov processes with discrete state spaces) the generalization as presented in Section 4 was not necessary. The Giry monad was introduced in (Eberl et al. 2015).

(Immler 2012) already proved that for each projective family the projective limits exist, using the Daniell-Kolmogorov extension theorem. This is clearly more general in the shape of the projective family, but Daniell-Kolmogorov is restricted to state spaces M_i which are Borel spaces of Polish spaces. This topological restriction of the state spaces would also restrict the compositionality of our Markov processes.

Most other formalizations of measure theory focus either on analysis, e.g. in PVS (Lester 2007) or in HOL-Light for the Flyspeck project (Hales et al. 2015), on traces of discrete spaces, e.g. in HOL (Hurd 2002; Liu et al. 2013), or on denotational semantics with discrete spaces (Hurd et al. 2005; Cock 2012; Audebaud and Paulin-Mohring 2009). None of these projects allows the combination of traces and continuous state spaces.

A formalization of Markov kernels for Isabelle/HOL was already done in (Backes et al. 2008; Berg 2013). However, it seams that some theorems requiring measure theory concepts were skipped using `sorry` (Berg 2013, p. 11).

3. Preliminaries

We will often use product sets $\Pi_{i\in I}\ F_i = \{f :: \alpha \Rightarrow \beta \mid \forall i \in I.\ f\ i \in F_i\}$, which have an explicit index set $I :: \alpha$ set and for each index $i \in I$ a set $F_i :: \beta$ set. As the set I is usually not the set of all elements of type α we sometimes restrict a function to a fixed value outside of I written $f|_I = (\lambda i.\ \textbf{if } i \in I \textbf{ then } f\ i \textbf{ else } \mathsf{undefined})$, where $\mathsf{undefined} :: \beta$ is an unspecified element. We write for the everywhere undefined function $\bot = (\lambda\ \text{-}\ .\ \mathsf{undefined})$.

When modelling the trace spaces of Markov models we have two different views on them: one is as a function $\omega :: \mathsf{nat} \Rightarrow \alpha$, the other one is as a co-datatype $\omega :: \alpha\ \mathsf{stream}$. Here streams are a co-inductively defined datatype as formalized in (Blanchette et al. 2014b). Streams have one constructor with two elements: a head element of type α and another stream as tail element. While both views are isomorphic, the stream view gives us a more list-like recursive behavior.

Streams are introduced using Isabelle's **codatatype**-command (Blanchette et al. 2014a):

$$\alpha\ \mathsf{stream} = \mathsf{SCons}\ \alpha\ (\alpha\ \mathsf{stream})$$

The **codatatype**-command defines the map function $\mathsf{map} :: (\alpha \Rightarrow \beta) \Rightarrow \alpha\ \mathsf{stream} \Rightarrow \beta\ \mathsf{stream}$. The function $\mathsf{to\_stream} :: (\mathsf{nat} \Rightarrow \alpha) \Rightarrow \alpha\ \mathsf{stream}$ converts sequences to streams, ω_n is the n-th element of ω, and $\mathsf{streams}\ S = \{\omega \mid \forall i.\ \omega_i \in S\}$.

Probability and Measure Theory The measure and probability theory used in this article was introduced in Isabelle/HOL in (Hölzl and Heller 2011), and has since then been continuously extended (Hölzl and Nipkow 2012c,b; Hölzl 2013; Avigad et al. 2016; Eberl et al. 2015; Gouezel 2015; Hölzl 2016). It is generic enough to be used for other applications than Markov processes, as demonstrated by (Gouezel 2015) and by (Avigad et al. 2016). In this section we give a short overview of the existing measure spaces. A more detailed overview of the theory is given in (Hölzl and Heller 2011; Hölzl 2013, 2016). A general introduction into measure theory is found in (Pollard 2002).

In Isabelle/HOL the type of measures $\alpha\ \mathsf{measure}$ contains measurable spaces on the type α, i.e. sets of mea-

surable sets, and measures on these sets. For each element $M :: \alpha$ measure, we have the following projection functions: space $M :: \alpha$ set is the carrier set, sets $M :: \alpha$ set set is the set of measurable sets, and measure $M :: \alpha$ set $\Rightarrow$ real is the measure itself. The measurable sets sets M form a σ-algebra on space M, i.e. sets M is a set of sets which contains the empty set, and which is closed under complement and countable union. The function measure M is a measure on sets M, i.e. it is a non-negative and countably-additive function and satisfies measure $M\ \emptyset = 0$.

We also use the type α measure to model measurable spaces, i.e. each measurable space has a measure associated to it. We abbreviate $M\ A$ to stand for measure $M\ A$. Note also that in this paper we are only working with probability measures, i.e. M (space M) $= 1$. We write set comprehensions modulo space M: $\{x \mid P\,x\} = \{x \in \text{space}\ M \mid P\,x\}$.

A predicate P holds almost-everywhere on M when the measure of its complement is 0:

$$AE\ x \text{ in } M.\ P\,x \longleftrightarrow \\ \exists N \in \text{sets}\, M.\ M\ N = 0 \wedge \{x \mid \neg\, P\,x\} \subseteq N.$$

The measurable space sigma $\Omega\ A$ is the smallest σ-algebra on the space Ω s.t. all sets in A are measurable. $\bigsqcup_{i\in I}\ M_i$ is the smallest σ-algebra S s.t. for all $i \in I$ the measurable sets of M_i are also measurable in S.[1]

We say a measurable space M is generated by G if $M =$ sigma $\Omega\ G$. If a measurable space is generated by G and G is closed under intersection and contains a countable cover of Ω, then the equality of two finite measures M and N on this space can be reduced to equality under G:

$$M = N \longleftrightarrow \forall A \in G.\ M\ A = N\ A.$$

The product space $M \times N$ is generated by the sets $A \times B$ for $A \in M$ and $B \in N$. The indexed product $\Pi_{i\in I}\ M_i$ is generated by the embedded sets $\{\omega \mid \forall j \in J.\ \omega\ j \in F_j\}$ where J is a finite subset of I and $F_j \in M_j$ for all $j \in J$.

Measurable functions Measurability is lifted from sets to functions:

$$\begin{aligned} &\text{-} \to_m \text{-} :: \alpha \text{ measure} \Rightarrow \beta \text{ measure} \Rightarrow (\alpha \Rightarrow \beta) \text{ set} \\ &M \to_m N = \{f \in \text{space}\ M \to \text{space}\ N \mid \\ &\qquad \forall A \in \text{sets}\ N.\ \{x \mid f\,x \in A\} \in \text{sets}\ M\} \end{aligned}$$

For $f \in M \to_m N$, we write that f is N-measurable on M, omitting either N or M when the reader can infer them from the context. The arrow notation already indicates it: measurability is similar to typing. In the rest of the paper we will often write the measurability statement of a function instead of its type. It is easy to derive the type from the measurability statement.

[1] We instantiate the complete lattice type class for α measure, with a lexicographic ordering where we first compare space, then sets, and finally measure. This allows us to use the usual lattice operators on measurable spaces and measures.

We write vsigma $\Omega\ M\ f$ to denote the smallest σ-algebra on Ω, s.t. f is M-measurable. For the extended non-negative real numbers ennreal $= [0, \infty]$, we define the Borel space borel $=$ sigma UNIV $\{S \mid \text{open}\ S\}$. All open sets, all closed sets, and intervals are Borel-measurable. From the definition follows that every continuous function is measurable. For many functions used in Isabelle/HOL we provide measurability rules, e.g. assuming f, g, and P are measurable then $f + g$, $f - g$, $f * g$, **if** P **then** f **else** g, fst, snd, etc. are measurable. Also $(\lambda\omega.\ \omega\ i) \in \left(\Pi_{i\in I}\ M_i\right) \to_m M_i$ if $i \in I$.

Each measure M gives rise to a non-negative Lebesgue integral $\int_x f\ x\ \mathrm{d}M$ for a Borel-measurable function f on M. For measurable functions we prove the usual theorems about the Lebesgue integral, i.e. that it is monotone and linear and that dominated and monotone convergence holds. We use the following notations

$$\int_{x\in A} f\ x\ \mathrm{d}M\ = \int_x \textbf{if}\ x \in A\ \textbf{then}\ f\ x\ \textbf{else}\ 0\ \mathrm{d}M \text{ and} \\ \int_{x|P\ x} f\ x\ \mathrm{d}M = \int_x \textbf{if}\ P\ x\ \textbf{then}\ f\ x\ \textbf{else}\ 0\ \mathrm{d}M\ .$$

Giry Monad An important construction mechanism for probabilistic programming language semantics is the *Giry monad* (Giry 1982), providing a monadic structure on probability measures. We introduce prob-algebra, a measurable space on the set of probability measures. This formalization is based on (Eberl et al. 2015) where prob-algebra is used for the denotational semantics of a simple programming language. Other formalizations of the Giry monad (Audebaud and Paulin-Mohring 2009; Hölzl et al. 2015) are restricted to discrete spaces.

In this paper the Giry monad is used to define Markov kernels and to express various equations on Markov chains and processes. Stating these rules in terms of measures, possible due to the Giry monad, gives less cluttered theorems and a straightforward method to derive the corresponding statements about integrals, measures, and the almost everywhere quantifier.

We define the measurable space prob-algebra K of probability measures on the measurable space K:

$$\begin{aligned} &\text{prob-algebra} :: \alpha \text{ measure} \Rightarrow \alpha \text{ measure measure} \\ &\text{prob-algebra}\ K = \textstyle\bigsqcup_{A\in \text{sets}\ K} \\ &\quad \text{vsigma}\ \{M \mid \text{sets}\ M = \text{sets}\ K \wedge \text{prob-space}\ M\} \\ &\qquad \text{borel}\ (\lambda M.\ M\ A) \end{aligned}$$

Then, for all measurable sets $A \in K$, the function $\lambda M.\ M\ A$ is measurable. Using prob-algebra we define the operators for the Giry monad.

$$\begin{array}{lcl} \text{map} & :: & (\alpha \Rightarrow \beta) \Rightarrow \alpha \text{ measure} \Rightarrow \beta \text{ measure} \\ \text{bind} & :: & \alpha \text{ measure} \Rightarrow (\alpha \Rightarrow \beta \text{ measure}) \Rightarrow \beta \text{ measure} \\ \text{return} & :: & \alpha \Rightarrow \alpha \text{ measure} \end{array}$$

With the notation $M \gg\!= f = \text{bind}\ M\ f$ and the equation map $f\ M = M \gg\!= (\lambda x.\ \text{return}\ (f\ x))$. In Isabelle there is also an explicit parameter to map $f\ M$ (`distr M N f` in

Isabelle) and return x (`return N x` in Isabelle) for a measurable space N. This is necessary as N can not be computed from f, x, or M. We do not show these measurable spaces since the user can easily derive them from the context and they would clutter the presentation.

We have the following measurability rules and defining equations for the monadic operators. Assuming the functions $f \in M \to_m N$, $A \in M \to_m \mathsf{prob\text{-}algebra}\ N$, $(\lambda(x,y).\ B\ x\ y) \in M \times N \to_m \mathsf{prob\text{-}algebra}\ L$ and a value $x \in M$. Also g is assumed to be Borel-measurable.

$$\begin{aligned}
&\mathsf{map}\ f \in \mathsf{prob\text{-}algebra}\ M \to_m \mathsf{prob\text{-}algebra}\ N\\
&\textstyle\int_x g\ x\ \mathsf{d}(\mathsf{map}\ f\ M) = \int_x g\ (f\ x)\ \mathsf{d}M\\
&\mathsf{return} \in M \to_m \mathsf{prob\text{-}algebra}\ M\\
&\textstyle\int_x g\ x\ \mathsf{d}(\mathsf{return}\ x) = g\ x\\
&(\lambda x.\ A\ x \gg\!= B\ x) \in M \to_m \mathsf{prob\text{-}algebra}\ L\\
&\textstyle\int_x f\ x\ \mathsf{d}(A \gg\!= B) = \int_y \left(\int_x g\ x\ \mathsf{d}(B\ y)\right)\ \mathsf{d}A
\end{aligned}$$

The measurability statement for bind looks quite complicated, it is however necessary to express that bind is measurable in both its parameters. There is no measurable space of measurable functions, so we state the measurability of B in an uncurried way using the product space $M \times N$.

Under similar conditions (i.e. all occurring functions are measurable, and all occurring values are in the corresponding space), we prove the monadic laws for bind and return:

$$\begin{aligned}
&\mathsf{bind}\ M\ \mathsf{return} &&= M\\
&\mathsf{bind}\ (\mathsf{return}\ x)\ f &&= f\ x\\
&\mathsf{bind}\ (\mathsf{bind}\ M\ f)\ g &&= \mathsf{bind}\ M\ (\lambda x.\ \mathsf{bind}\ (f\ x)\ g)
\end{aligned}$$

Stream Spaces We introduce the canonical measurable space on streams, for which all (co)recursively defined functions are measurable. This is isomorphic to the infinite product measure $\Pi_{n::\mathsf{nat}}\ M$, so we define it as an embedding:

$$\begin{aligned}
&\mathsf{stream\text{-}space} &&:: \alpha\ \mathsf{measure} \Rightarrow \alpha\ \mathsf{stream\ measure}\\
&\mathsf{stream\text{-}space}\ M &&=\\
&\quad \mathsf{vsigma}\ (\mathsf{streams}\ (\mathsf{space}\ M))\ (\lambda\omega\ i.\ \omega_i)\ (\Pi_{n::\mathsf{nat}}\ M)
\end{aligned}$$

(The definition in Isabelle/HOL is more complicated, as it assigns also a canonical measure to stream-space M.) By this definition, the measurable space of stream-space is the smallest measurable space s.t. projection of elements is measurable. Hence a function which makes all projections measurable is also measurable:

$$(\lambda\omega.\ \omega_i) \in \mathsf{stream\text{-}space}\ M \to_m M$$

$$\frac{\forall i.\ (\lambda x.\ (f\ x)_i) \in N \to_m M}{f \in N \to_m \mathsf{stream\text{-}space}\ M}$$

The measurability rules of map, and to_stream are easily derived by these two rules.

Since the stream space is defined as a mapping of the product space, we know that two measures on the stream space are equal if they are equal for each cylinder set $C_{n,A} = \{\omega | \forall i < n.\ \omega_i \in A_i\}$ with $A_i \in M$ forall $i < n$. With this we derive the following coinduction rule for measures on stream spaces:

Lemma 1. *Given two measures A, B with* sets A = sets B = stream-space M*, and a relation R with $R\ A\ B$. Also given that R is closed under the following assumption: For each A and B with $R\ A\ B$ exists a $K \in$ prob-algebra M, and $A', B' \in M \to_m$ prob-algebra (stream-space M), such that:*

$$\begin{aligned}
&AE\ y\ \mathsf{in}\ K.\ R\ (A'\ y)\ (B'\ y) \vee A'\ y = B'\ y,\\
&A = \textbf{\textit{do}}\ \{y \leftarrow K; \omega \leftarrow A'\ y; \mathsf{return}\ (y{\cdot}\omega)\}, \textit{and}\\
&B = \textbf{\textit{do}}\ \{y \leftarrow K; \omega \leftarrow B'\ y; \mathsf{return}\ (y{\cdot}\omega)\}.
\end{aligned}$$

Then $A = B$.

4. Discrete-Time Markov Processes

In this section we show how to construct discrete-time Markov processes from a Markov kernel. This construction fulfills an iteration rule, a bisimulation property and the strong Markov property. We do this with the following steps:

- We formalize the extension theorem on probability measures by Ionescu-Tulcea following the presentation in (Pollard 2002). This theorem allow us to construct a probability measure for the direct limit of a sequence of random variables X where each X_i depends on all previous X_j with $j < i$. This dependency is described by a Markov kernel.
- Using the Ionescu-Tulcea extension theorem we construct the discrete-time Markov process into sequences. We show that this is a Markov kernel in the initial state.
- Finally, we lift the discrete-time Markov process from sequences to streams and prove the strong Markov property using the coinduction property for measures on streams. For the latter we introduce the concept of stopping times.

Extension theorem by Ionescu-Tulcea We assume a sequence of measurable spaces M and a sequence of Markov kernels $K\ n \in (\Pi_{i<n}\ M_i) \to_m \mathsf{prob\text{-}algebra}\ M_n$. The measurable spaces M_n can be seen as state spaces and $K\ n$ as transition functions depending on the n previous states. We define the function $\mathsf{C}\ n\ m$ to iterate K from n to $n+m$, i.e. given n starting states it computes the next m steps and then returns a combining distribution on $n+m$ states:

$$\begin{aligned}
&\mathsf{eK}\ n \in (\Pi_{i<n}\ M_i) \to_m \mathsf{prob\text{-}algebra}\ (\Pi_{i<n+1}\ M_i)\\
&\mathsf{eK}\ n\ \omega &&= \mathsf{map}\ (\lambda x.\ \omega(n := x))\ (K\ n\ \omega)\\
&\mathsf{C}\ n\ m \in (\Pi_{i<n}\ M_i) \to_m \mathsf{prob\text{-}algebra}\ (\Pi_{i<n+m}\ M_i)\\
&\mathsf{C}\ n\ 0\ \omega &&= \mathsf{return}\ \omega\\
&\mathsf{C}\ n\ (m+1)\ \omega &&= \mathsf{C}\ n\ m\ \omega \gg\!= \mathsf{eK}\ (n+m)
\end{aligned}$$

From C we construct C', which is C restricted to a finite index set I ($\mathsf{Max}\ I$ is the maximal number in I, $\mathsf{Max}\ \emptyset = 0$):

$$\begin{aligned}
&\mathsf{C'}\ I &&\in \mathsf{prob\text{-}algebra}\ (\Pi_{i\in I}\ M_i)\\
&\mathsf{C'}\ I &&= \mathsf{map}\ (\lambda\omega.\ \ \omega|_I)\ (C\ 0\ (\mathsf{Max}\ I + 1)\ \bot)
\end{aligned}$$

The measures C' form a projective family in I, i.e. for all finite sets I and $I' \subseteq I$

$$\mathsf{map}\ (\lambda\omega.\ \omega|_{I'})\ (\mathsf{C}'\ I) = \mathsf{C}'\ I'\ .$$

We show that C' has a projective limit: that there exists a measure L, s.t. L can be projected on $\mathsf{C}'\ I$ for all finite sets I.

Theorem 1 (Extension theorem by Ionescu-Tulcea)**.** *There exists a probability measure* $\mathsf{IT}\ M\ K$ *which is the projective limit of* C'*, i.e.* $\mathsf{IT}\ M\ K \in \mathsf{prob\text{-}algebra}\ (\Pi_i\ M_i)$ *and for all finite* I *we have*

$$\mathsf{map}\ (\lambda\omega.\ \omega|_I)\ (\mathsf{IT}\ M\ K) = \mathsf{C}'\ I\ .$$

We call this limit $\mathsf{IT}\ M\ K$.

Proof. In Isabelle/HOL we formalized the proof as presented in (Pollard 2002) using Caratheodory's extension theorem. Given a sequence of finite subsets I_n and a sequence of measurable sets $X_n \in \Pi_{i\in I_n}\ M_i$, where the sequence of embeddings $\{\omega \mid \omega|_{I_n} \in X_n\}$ is decreasing and $\bigsqcap_n \mathsf{C}'\ I_n\ X_n > 0$. To finish the proof, we need to construct a ω', s.t. $\omega'|_{I_n} \in X_n$ for all n. Without loss of generality, we assume that $I_n = \{0, \ldots, n-1\}$, hence $\bigsqcap_n \mathsf{C}\ 0\ n \perp X_n > 0$.

Given j, $\omega \in \Pi_{i<j}\ M_i$ and $\bigsqcap_n \mathsf{C}\ j\ n\ \omega\ X_{j+n} > 0$ we calculate:

$$\begin{aligned} 0 &< \textstyle\bigsqcap_n \mathsf{C}\ j\ n\ \omega\ X_{j+n} \\ &\leq \textstyle\bigsqcap_n \mathsf{C}\ j\ (n+1)\ \omega\ X_{(j+1)+n} \\ &= \textstyle\bigsqcap_n \int_x \mathsf{C}\ (j+1)\ n\ (\omega(j := x))\ X_{(j+1)+n}\ d\mathsf{K}\ n\ \omega \\ &= \textstyle\int_x \bigsqcap_n \mathsf{C}\ (j+1)\ n\ (\omega(j := x))\ X_{(j+1)+n}\ d\mathsf{K}\ n\ \omega \end{aligned}$$

Hence there exists a x, s.t. $\omega(j := x) \in X_{j+1}$ and

$$\bigsqcap_n \mathsf{C}\ (j+1)\ n\ \omega\ X_{(j+1)+n} > 0\ .$$

Using dependent choice we construct the sequence ω', s.t. $\omega'|_{I_n} \in X_n$ for all n. □

Given a constant Markov kernel K, we get an infinite product space with the natural numbers as index set. We can extend this to arbitrary index sets: we only need to show countable additivity of the product measure on sets embedded from finite index sets. This allows us to prove the following corollary:

Corollary 1. *Given a (possibly uncountable) infinite set* I *and probability measures* M_i *for* $i \in I$*, the product probability measure* $\Pi_{i\in I}\ M_i$ *exists.*

So we not only have a measurable space $\Pi_{i\in I}\ M_i$, but we can now also assign a canonical measure. The infinite product measure was already formalized in (Hölzl 2013), but used an ad-hoc proof. The proof of Theorem 1 (including its constructions) is around $\sim$ 340 lines. The ad-hoc proof of Corollary 1 had $\sim$ 300 lines before and was replace by a $\sim$ 90 lines proof.

Construct Markov processes on sequences From Ionescu-Tulcea we construct the Markov process on sequences $\mathsf{proc}_{\mathsf{seq}}$. In this section we assume a single state space M, and a Markov kernel $K \in M \to_m \mathsf{prob\text{-}algebra}\ M$. The central definition is $\mathsf{proc}_{\mathsf{seq}}$:

$$\begin{aligned} &\mathsf{proc}_{\mathsf{seq}} \in M \to_m \mathsf{prob\text{-}algebra}\ (\Pi_i\ M) \\ &\mathsf{proc}_{\mathsf{seq}}\ x = \mathsf{IT}\ (\lambda\ \text{-}\ .\ M)\ (\lambda n\ \omega.\ K\ ((x{\cdot}\omega)_n)) \end{aligned}$$

Here $\Pi_i\ M$ is the measurable space on all sequences into M, but M does not depend on the position i. In Isabelle/HOL, we first write the equation and prove afterwards the measurability condition. We prove this condition to show that $\mathsf{proc}_{\mathsf{seq}}$ is well-defined:

Lemma 2 ($\mathsf{proc}_{\mathsf{seq}}$ is a Markov kernel)**.**

$$\mathsf{proc}_{\mathsf{seq}} \in M \to_m \mathsf{prob\text{-}algebra}\ (\Pi_i\ M)$$

Proof. The measurable space $\Pi_i\ M$ is generated by all finite projections. Hence it is enough to show that, given a finite set I and a measurable set $X \in \Pi_{i\in I}\ M$, then

$$(\lambda x.\ \mathsf{proc}_{\mathsf{seq}}\ x\ \{\omega \mid \omega|_I \in X\}) \in M \to_m \mathsf{borel}\ . \qquad (5)$$

Without loss of generality we assume that there exists a n, s.t. $I = \{0, \ldots, n\}$. Then by applying Theorem 1 to Eq. (5), we have

$$(\lambda x.\ \mathsf{C}_x\ 0\ n\ X) \in M \to_m \mathsf{borel}\ . \qquad (6)$$

Note that C has now the hidden parameters $\lambda\ \text{-}\ .\ M$ and $\lambda n\ \omega.\ K\ ((x{\cdot}\omega)_n)$, so we write it with a subscripted x. Eq. (6) is finally proved by induction on n, and the measurability of map, bind, and return on measures. □

The measurability of $\mathsf{proc}_{\mathsf{seq}}$ is necessary to show our second important characterizing property of $\mathsf{proc}_{\mathsf{seq}}$: how to compute with it. The intuitive understanding is: the process starts in a state x, calls $K\ x$ to randomly choose another state y, and then continue by calling itself with y, while generating an infinite stream $(y{\cdot}\omega)$. This is now concisely expressed using the Giry monad:

Theorem 2 (Iteration rule for $\mathsf{proc}_{\mathsf{seq}}$)**.** *For all* $x \in M$ *holds*

$$\mathsf{proc}_{\mathsf{seq}}\ x = \textbf{\textit{do}}\ \{y \leftarrow K\ x; \omega \leftarrow \mathsf{proc}_{\mathsf{seq}}\ y; \mathsf{return}\ (y{\cdot}\omega)\}\ .$$

Proof. We prove the equality by showing that both measures are equal on the sets generating $\Pi_i\ M$. As generating sets we choose the sets $\Pi_{i\in I}\ F\ i$ for all finite sets I and $F\ i \in M_i$ for $i \in I$. This is a valid set of sets to choose as it generates $\Pi_i\ M$ and is closed under intersection. Without loss of generality we assume there exists an n, s.t. $I = \{0, \ldots, n+1\}$. We prove by induction on n that $\mathsf{C}_x\ 1\ n\ (y{\cdot}\perp) =$

$\mathsf{C}_y\ 0\ n\ \bot$. With this we compute:

$$\begin{aligned}
&\mathsf{proc}_{\mathsf{seq}}\ x\ (\Pi_{i<n+1}\ F\ i)\\
&= \mathsf{C}_x\ 0\ (n+1)\ \bot\ (\Pi_{i<n+1}\ F\ i)\\
&= \textstyle\int_{y\in F\ 0} \mathsf{C}_x\ 1\ n\ (y{\cdot}\bot)\ (\Pi_{i<n}\ F\ (i+1))\,\mathsf{d}K\ x\\
&= \textstyle\int_{y\in F\ 0} \mathsf{C}_y\ 0\ n\ \bot\ (\Pi_{i<n}\ F\ (i+1))\,\mathsf{d}K\ x\\
&= \textstyle\int_{y\in F\ 0} \mathsf{proc}_{\mathsf{seq}}\ y\ (\Pi_{i<n}\ F\ (i+1))\,\mathsf{d}K\ x\\
&= \textstyle\int_{y} \big(\mathsf{map}\ (\lambda\omega.\ y{\cdot}\omega)\ (\mathsf{proc}_{\mathsf{seq}}\ y)\big)\ \ (\Pi_{i<n+1}\ F\ i)\,\mathsf{d}K\ x\\
&= \mathbf{do}\ \big\{y \leftarrow K\ x; \omega \leftarrow \mathsf{proc}_{\mathsf{seq}}\ y; \mathsf{return}\ (y{\cdot}\omega)\big\}\\
&\quad (\Pi_{i<n+1}\ F\ i)
\end{aligned}$$

□

Is $\mathsf{proc}_{\mathsf{seq}}$ unique? Are there other Markov processes on $\Pi_i\ M$ with the iteration rule? It turns out that there are no other Markov processes, that $\mathsf{proc}_{\mathsf{seq}}$ is unique on $\Pi_i\ M$.

But, before we prove uniqueness, we want to change the space we operate in. In the proof of the extension theorem for Ionescu-Tulcea, and the proofs of the measurability and iteration rules for $\mathsf{proc}_{\mathsf{seq}}$ we were always reducing to cylinder sets, i.e. embeddings X from an index set $\{0,\ldots,n\}$. We want to make this more explicit, by using the space on streams.

Discrete-time Markov processes on streams We define the process construction on streams as a lifting from processes on sequences:

$$\begin{aligned}
&\mathsf{proc}_{\mathsf{stream}} \in M \to_m \mathsf{prob\text{-}algebra}\ (\mathsf{stream\text{-}space}\ M)\\
&\mathsf{proc}_{\mathsf{stream}}\ x = \mathsf{map\ to_stream}\ (\mathsf{proc}_{\mathsf{seq}}\ x)
\end{aligned}$$

The iteration rule looks exactly the same (remember that in this paper the $x{\cdot}\omega$ notation is used for both types, for $\mathsf{nat}{\Rightarrow}\alpha$ and $\alpha\ \mathsf{stream}$).

$$\begin{aligned}
&\mathsf{proc}_{\mathsf{stream}}\ x =\\
&\quad \mathbf{do}\ \big\{y \leftarrow K\ x; \omega \leftarrow \mathsf{proc}_{\mathsf{stream}}\ y; \mathsf{return}\ (y{\cdot}\omega)\big\}\ .
\end{aligned}$$

Using the coinduction rule for stream spaces we also show uniqueness of this construction.

Theorem 3 (Coinduction rule for $\mathsf{proc}_{\mathsf{stream}}$)**.** *Given a relation $R :: \alpha \to \alpha\ \mathsf{stream\ measure} \to \mathsf{bool}$, a state $x \in M$ and a measure N, where $R\ x\ N$ holds. Also assume that for each x and N for which $R\ x\ N$ holds there exists an $N' \in M \to_m \mathsf{prob\text{-}algebra}(\mathsf{stream\text{-}space}\ M)$, such that*

$$AE\ y\ \mathsf{in}\ K\ x.\ R\ y\ (N'\ y) \vee \mathsf{proc}_{\mathsf{stream}}\ y = N'\ y$$

and

$$N = \mathbf{do}\ \{y \leftarrow K\ x; \omega \leftarrow N'\ y; \mathsf{return}\ (y{\cdot}\omega)\}\ .$$

Under these assumptions $\mathsf{proc}_{\mathsf{stream}}\ x = N$ *holds.*

The structure of this rule allows that it can be applied using Isabelle's `coinduction` method (Blanchette et al. 2014a). When applying `coinduction` on an equality statement it will collect all the occurring variables and assumptions and automatically create the relation R.

The strong Markov property The defining property of Markov processes is that their future behavior only depends on the current state, i.e. that they do not depend on the history before it. This is the *Markov property*, expressed in the iteration rule which states that the probability of a Markov process can be split up by first choosing a y from $K\ x$ and then continuing in y, ignoring the x before. An extension of this is the *strong Markov property*, which tells us that we can stop the Markov process at an arbitrary time, just remember the state at this time and continue with that state. To formalize the concept of *arbitrary time* we need to introduce the concept of *stopping times*, measurable functions from our Markov process into time values.

The following definition of stopping time and filtration is general in the time domain τ, a linearly ordered second-countable[2] topology. However, the only instance of τ used here are the extended natural numbers $\mathsf{enat} = \mathsf{nat} \cup \{\infty\}$.

A function $F :: \tau{\Rightarrow}\alpha\ \mathsf{measure}$ is a *filtration* F on a space Ω if F is monotone and the space of $F\ t$ is Ω for all t, we write $\mathsf{filtration}\ \Omega\ F$. Intuitively, the measurable space $F\ t$ describes the observations we were able to make until time t.

For a *stopping time* T with regard to a filtration F the time value $T\ \omega$ does only depend on the observations up to this time $T\ \omega$. This is easily expressed using measurability with regard to F:

$$\mathsf{stopping\_time}\ F\ T \longleftrightarrow \big(\forall t.\ \big\{\omega \mid T\ \omega \le t\big\} \in F\ t\big)$$

Typical stopping times are constant functions, minimum and maximum of other stopping times, or the first time a set of states is reached.

For our discrete-time Markov processes we define the filtration on streams which allows to observe the states up to time t.

$$\mathsf{F}_{\mathsf{stream}}\ M\ n = \bigsqcup_{i::\mathsf{nat}<n} \mathsf{vsigma}\ (\mathsf{streams}\ M)\ (\lambda\omega.\ \omega_i)\ M$$

Here n is an extended natural number and the bound variable i is a natural number. Hence, $\mathsf{F}_{\mathsf{stream}}\ M\ \infty$ will be the entire trace space of our process. The filtration $\mathsf{F}_{\mathsf{stream}}\ M$ is also called the filtration adapted to streams, as it is adapted to the sequence of random variables $\lambda\omega.\ \omega_i$.

Before we prove the strong Markov property we show three properties of a stopping time T with regard to the filtration $\mathsf{F}_{\mathsf{stream}}\ M$. In the following we assume a state $x \in M$ and a stream $\omega \in \mathsf{streams}\ M$. First, the set of states x for which $T\ (x{\cdot}\omega)$ returns 0 is measurable:

$$\{x \mid \forall\omega \in \mathsf{streams}\ M.\ T\ (x{\cdot}\omega) = 0\} \in M \qquad (7)$$

Also, if $T(x{\cdot}\omega) > 0$ holds, then so does it for all other ω':

$$\forall\omega, \omega' \in \mathsf{streams}\ M.\ T(x{\cdot}\omega) > 0 \longrightarrow T(x{\cdot}\omega') > 0 \qquad (8)$$

[2] A second countable topology is a topology generated by a countable base B, i.e. each open set is the union of a subset of B.

And finally, when $T(x{\cdot}\omega) > 0$ holds, then a "shifted" T is also a stopping time:

$$\mathsf{stopping\_time}\ (\mathsf{F}_{\mathsf{stream}}\ M)\ (\lambda\omega.\ T(x{\cdot}\omega) - 1) \tag{9}$$

Now we state and prove the strong Markov property.

Theorem 4 (Strong Markov property)**.** *Given a stopping time T adapted to $\mathsf{F}_{\mathsf{stream}}\ M$ and a state $x \in M$. Then the following equation holds:*

$$\begin{aligned}
&\mathsf{proc}_{\mathsf{stream}}\ x = \textbf{\textit{do}}\ \{\\
&\quad \omega \leftarrow \mathsf{proc}_{\mathsf{stream}}\ x\\
&\quad \textbf{\textit{case}}\ T\ \omega\ \textbf{\textit{of}}\\
&\quad\quad \mid \infty \quad\quad \Rightarrow \mathsf{return}\ \omega\\
&\quad\quad \mid n :: \mathsf{nat} \Rightarrow \omega' \leftarrow \mathsf{proc}_{\mathsf{stream}}\ \omega_n\\
&\quad\quad\quad\quad\quad\quad \mathsf{return}\ (\omega \cdot_{n+1} \omega')\\
&\}
\end{aligned} \tag{10}$$

Where $\omega \cdot_n \omega'$ is the stream consisting of the first n elements of ω and then continuing with ω'.

Proof. Proof by coinduction with Theorem 3, where the relation R describes that $x \in M$ and that there exists a stopping time T. In the rest of the proof we use $P\ T\ x$ for the right-hand side of Eq. (10).

Now, for each $x \in M$ and stopping time T, we define

$$T'\ y\ \omega = T\ (y{\cdot}\omega) - 1$$

$$N'\ y\ = \begin{cases} P\ (T'\ y)\ y & \textbf{if}\ \ \forall\omega \in \mathsf{streams}\ M.\ T\ (y{\cdot}\omega) > 0\\ \mathsf{proc}_{\mathsf{stream}}\ y & \textbf{otherwise}\ . \end{cases}$$

We know by Eq. (9) that T' is a stopping time in the context of N'. Then N' is a Markov kernel by Eq. (7), the measurability of $\mathsf{proc}_{\mathsf{stream}}$, and the measurability of the monadic operators. From the definition of N' follows immediately:

$$AE\ y\ \mathsf{in}\ K\ x.\ N'\ y = P\ (T'\ y)\ y \vee \mathsf{proc}_{\mathsf{stream}}\ y = N'\ y$$

Finally, to finish the coinduction proof we need to show:

$$P\ T\ x = \textbf{do}\ \{y \leftarrow K\ x; \omega \leftarrow N'\ y; \mathsf{return}\ (y{\cdot}\omega)\}\ .$$

Which is solved by the iteration rule on $\mathsf{proc}_{\mathsf{stream}}\ x$, the case distinction in N', Eq. (8), as well as rewriting with the monad laws. □

5. Continuous-Time Markov Chains

Before we describe continuous-time Markov chains as stochastic processes, we introduce transition rates. In this section we assume continuous-time Markov chains are specified as a function R of *transition rates*

$$R :: \alpha \Rightarrow \alpha \Rightarrow \mathsf{real}$$

with the assumptions that it

(R1) is non-negative, $0 \leq R\ x\ y$ for all x and y,

(R2) has a zero diagonal, $R\ x\ x = 0$ for all x, and

(R3) has finite and positive columns, $0 < \sum_y R\ x\ y < \infty$ for all x.

$R\ \ x\ \ y$ is the transition rate to go from state x to state y. There is no restriction on the type of α, it does not need to be countable or even finite. From **(R3)**, we derive that $\{y \mid R\ x\ y \neq 0\}$ is countable. Note that our usage of transition rates is different from the often used Q-matrix, where the diagonal is the negative of the entire column:

$$Q\ x\ x = -\sum_y R\ x\ y \qquad Q\ x\ y = R\ x\ y \text{ if } x \neq y\ .$$

A state y is *enabled* in x if $R\ x\ y > 0$, we write $y \in I\ x$. A state y is accessible from x if it is reachable through finitely many enabled states. From **(R3)** it follows that the set of enabled states $I\ x$ is countable. The *escape rate* is the sum over a column: $E\ x = \sum_y R\ x\ y$. This allows us to define a discrete measure J (i.e. a probability mass function) describing the jump behaviour of our Markov chain. We define the state space S to be the discrete measurable space on α, i.e. in S all sets are measurable. We define $J\ x$ by its behaviour on a single element y:

$$\begin{aligned} &J \in S \rightarrow_m \mathsf{prob\text{-}algebra}\ S\\ &J\ x\ \{y\} = R\ x\ y / E\ x \end{aligned}$$

With this we define kernel K for our hold & jump process modelling the Markov chain. The kernel operates on a pair of current time and state, and returns a distribution of the future time and state. The difference between the current time and the future time is exponentially distributed, the state is chosen using J. The time is modelled as absolute time, not as holding time, so we need to map over the exponential distribution.

$$\begin{aligned} &K \in (\mathsf{borel} \times S) \rightarrow_m \mathsf{prob\text{-}algebra}\ (\mathsf{borel} \times S)\\ &K\ (t, x) = (\mathsf{map}\ (+t)\ (\mathsf{Exp}\ (E\ x))) \times J\ x \end{aligned} \tag{11}$$

First, it is important to note that the Markov chain is invariant under the time t with which it starts. If we translate the start time, the time stored in the traces is also shifted:

Corollary 2. *Invariant under time translation*

$$\begin{aligned} &\mathsf{proc}_{\mathsf{stream}}\ (t + t', x) =\\ &\quad \mathsf{map}\ (\mathsf{map}\ (\lambda(t, x).\ (t + t', x)))\ (\mathsf{proc}_{\mathsf{stream}}\ (t, x)) \end{aligned}$$

Proof. Follows by Theorem 3 □

A second question is: is this the right model? Is it okay to do one step using the escape time? To answer this question with yes we show an alternative model for the kernel K based on a choice from independent runs.

The kernel K as choice from independent runs We assume that for each state x we look at the distribution of independent runs for each enabled state $y \in I\ x$. An independent run is described by the time the event y happens, which is exponentially distributed with rate $R\ x\ y$. The product over all these times describes the independent behaviour:

$$\mathsf{parallel}\ x = \Pi_{y \in Ix}\ \mathsf{Exp}\ (R\ x\ y)$$

Our goal is to choose from a sample p of $\mathsf{parallel}\ x$ the first event, i.e. the y s.t. $p\ y$ is the shortest time:

$$\mathsf{first}\ x\ p\ y = y \in I\ x \wedge (\forall y' \in I\ x - \{y\}.\ p\ y < p\ y')$$

The definition already ensures that at most one y satisfies $\mathsf{first}\ x\ p\ y$. To use this definition we will need to show that such a y exists with probability 1.

Lemma 3 (Parallel choice up to time)**.** *Assume $y \in I\ x$ and a time t, then the probability that y happens first after time t is the same as choosing t exponentially and independently choosing y:*

$$\mathsf{parallel}\ x\ \{p \mid t \leq p\ y \wedge \mathsf{first}\ x\ p\ y\} = \\ \mathsf{Exp}\ (E\ x)\ [t, \infty) * J\ x\ \{y\}$$

Proof. The central idea of this proof is, that a countable product of exponential distributions indexed by S and parameterized in r equals an exponential distribution with parameter $\sum_{x \in S} r\ x$. We provide the following statement in S, r and t:

$$\left(\Pi_{x \in S}\ \mathsf{Exp}\ (r\ x)\right)\ \left(\Pi_{x \in S}\ [t, \infty)\right) = e^{-t * \sum_{x \in S} r\ x}$$

Now, we prove the theorem by calculation (we abbreviate $P' = \Pi_{y' \in I\ x - \{y\}}\ \mathsf{Exp}\ (R\ x\ y')$):

$$\begin{aligned}
&\mathsf{parallel}\ x\ \{p \mid t \leq p\ y \wedge \mathsf{first}\ x\ p\ y\} \\
&= \textstyle\int_{t' \geq t} P'\ \{p \mid \forall y' \in I\ x - \{y\}.\ t' \leq p\ y'\} \mathsf{dExp}\ (R\ x\ y) \\
&= \textstyle\int_{t' \geq t} P'\ \left(\Pi_{y' \in I\ x - \{y\}}\ [t', \infty)\right) \mathsf{dExp}\ (R\ x\ y) \\
&= \textstyle\int_{t' \geq t} e^{-t' * (E\ x - R\ x\ y)} \mathsf{dExp}\ (R\ x\ y) \\
&= (R\ x\ y)/(E\ x) * \mathsf{Exp}\ (E\ x)\ [t, \infty) \\
&= J\ x\ \{y\} * \mathsf{Exp}\ (E\ x)\ [t, \infty)
\end{aligned}$$

□

From this we derive that with probability 1 there is always exactly one y which happens first. We instantiate Lemma 3 with time 0 and sum up over all $y \in I\ x$.

Lemma 4. *AE p in* $\mathsf{parallel}\ x.\ \exists y \in I\ x.\ \mathsf{first}\ x\ p\ y$

With this lemma, the following statement is well defined: $\mathsf{The}\ (\mathsf{first}\ x\ p)$ is the unique y, s.t. $\mathsf{first}\ x\ p\ y$ holds.

Theorem 5 (K as independent runs)**.**

$$\begin{aligned}
K\ (t, x) = \textbf{\textit{do}}\ \{ \\
\quad p \leftarrow \mathsf{parallel}\ x \\
\quad y := \mathsf{The}\ (\mathsf{first}\ x\ p) \\
\quad \mathsf{return}\ (t + p\ y, y) \\
\}
\end{aligned}$$

Proof. Proof by uniqueness of measures on the generating sets $[t, \infty) \times \{y\}$ for all $y \in I\ x$. To cover the entire space we also need to add all $[t, \infty) \times A$ for $A \cap I\ x = \emptyset$ to the generating sets, obviously these sets have a null measure on both sides. The proof is finished with Lemma 3. □

Markov chain as stochastic process on* real *time In this section we assume the Markov kernel K as defined in Eq. (11) and $\mathsf{proc}_{\mathsf{stream}}$ instantiated by this K.

Usually, stochastic processes are formalized as a family of random variables X indexed by time t, s.t. $X_t \in P \to_m M$ where P is the ambient probability space in which the stochastic process lives and M is the state space. For this we introduce trace using Isabelle's **partial-function** package:

$$\mathsf{trace}\ x\ ((t, y) \cdot \omega)\ t' = \begin{cases} \mathsf{trace}\ y\ \omega\ t' & \textbf{if}\ \ t' \leq t \\ x & \textbf{otherwise} \end{cases}$$

Here x is the state which will be returned before the time t in the first element of ω. There is no implicit assumption that the time values in ω are strictly increasing. But a state value is only returned if a time strictly after t' is found. This results in a right-continuous step function, i.e. trace is constant on intervals of the form $[t, t')$.

A problem with this definition is that continuous-time Markov chains on infinite state spaces can *explode*, i.e. they show a Zeno effect where with a positive probability the times in a trace converge against a finite limit. This would require to encode an *explosion* state in M, or as happens by the definition with **partial-function** the arbitrary value undefined. However, it will turn out that in most cases we avoid this through other means.

Lemma 5. *Given a trace ω, a state x and a time t, if i is the first point in ω, s.t. $t < \mathsf{fst}\ (\omega_i)$ then*

$$\mathsf{trace}\ x\ \omega\ t = (x \cdot \mathsf{map}\ \mathsf{snd}\ \omega)_i$$

otherwise if no such i exists $\mathsf{trace}\ x\ \omega\ t = \mathsf{undefined}$.

Similar to discrete-time Markov processes, we also want to show the Markov property for continuous-time Markov chains. For this we need a way to merge two traces of time-state pairs at a specific time. Since the time stored in the trace is absolute we do not need to change it, we just need to find the point when to switch to the other stream.

$$\mathsf{merge}\ ((t, x) \cdot \omega)\ t'\ \omega' = \begin{cases} (t, x) \cdot \mathsf{merge}\ \omega\ t'\ \omega' & \textbf{if}\ t \leq t' \\ \omega' & \textbf{otherwise} \end{cases}$$

Theorem 6 (Markov property for continuous-time Markov chains)**.** *Given a starting state x at time t and a future time $t'' \geq t$, the following equation holds:*

$$\begin{aligned}
&\mathsf{proc}_{\mathsf{stream}}\ (t, x) = \textbf{\textit{do}}\{ \\
&\quad \omega \leftarrow \mathsf{proc}_{\mathsf{stream}}\ (t, x) \\
&\quad \omega' \leftarrow \mathsf{proc}_{\mathsf{stream}}\ (t'', \mathsf{trace}\ (t, x)\ \omega\ t'') \\
&\quad \mathsf{return}\ (\mathsf{merge}\ \omega\ t''\ \omega') \\
&\}
\end{aligned} \tag{12}$$

What happens if t'' is beyond the explosion time of ω? In this case $\mathsf{merge}\ \omega\ t''\ \omega' = \omega$ and we can ignore ω'. Note that this works as $\mathsf{proc}_{\mathsf{stream}}\ (t'', \ldots)$ is a probability space.

Proof. Proof by coinduction with Theorem 3, where the relation R describes that there is a t', s.t. $t' \geq t$. For each x, t, and t'' with $t'' \geq t$, we define N':

$$N' \in S \to_m \mathsf{prob\text{-}algebra}\ (\mathsf{stream\text{-}space}\ S)$$
$$N'\ (t', x') = \begin{cases} L\ t'\ x' & \textbf{if}\ t' \leq t'' \\ \mathsf{proc}_{\mathsf{stream}}\ (t', x') & \textbf{otherwise} \end{cases}$$

The probability measure $L\ t\ x$ is the abbreviation for the right-hand side of Eq. (12). For N' we show

$$AE\ (t', x')\ \mathsf{in}\ K\ (t, x).\ N'\ (t', x') = L\ (t', x') \vee N'\ (t', x') = \mathsf{proc}_{\mathsf{stream}}\ (t', x') \tag{13}$$

and

$$L\ t\ x = \textbf{do}\ \{tx \leftarrow K\ (t, x); \omega \leftarrow N'\ tx; \mathsf{return}\ (tx{\cdot}\omega)\}\ . \tag{14}$$

The measurability of N' and Eq. (13) are true by definition.

We introduce the abbreviation C, such that $L\ t\ x = K\ (t, x) \gg\!= C$ (by the iteration rule):

$$\begin{aligned} &C\ (t', x') = \textbf{do}\ \{ \\ &\quad \omega \leftarrow \mathsf{proc}_{\mathsf{stream}}\ (t', x') \\ &\quad \omega' \leftarrow \mathsf{proc}_{\mathsf{stream}}\ (t'', \mathsf{trace}\ x\ ((t', x'){\cdot}\omega)\ t'') \\ &\quad \mathsf{return}\ (\mathsf{merge}\ ((t', x'){\cdot}\omega)\ t''\ \omega') \\ &\} \end{aligned}$$

We show that Eq. (14) holds for all measurable sets A:

$$\begin{aligned} &(L\ t\ x)\ A \\ =\ &(K\ (t, x) \gg\!= C)\ A \\ =\ &\textstyle\int_{(t', x')|t' \leq t''} (C\ (t', x'))A\ \mathsf{d}K\ (t, x) + \\ &\textstyle\int_{(t', x')|t'' < t'} (C\ (t', x'))A\ \mathsf{d}K\ (t, x) \\ =\ &\textstyle\int_{(t', x')|t' \leq t''} \mathsf{map}\ ((t', x'){\cdot})\ (N'\ (t', x'))A\ \mathsf{d}K\ (t, x) + \\ &\textstyle\int_{(t', x')|t'' < t'} \mathsf{proc}_{\mathsf{stream}}\ (t', x')A\ \mathsf{d}K\ (t, x) \end{aligned}$$

The first integral, where $t' \leq t''$, $C\ (t', x')$ is already finished, we just need to show that the second integrand has the same form. For this we use that the exponential distribution is memory-less.

$$\begin{aligned} &\textstyle\int_{(t', x')|t'' < t'} (C\ (t', x'))A\ \mathsf{d}K\ (t, x) \\ =\ &\textstyle\int_{(t', x')|t'' < t'} (\mathsf{proc}_{\mathsf{stream}}\ (t'', x))A\ \mathsf{d}K\ (t, x) \\ =\ &\mathsf{proc}_{\mathsf{stream}}\ (t'', x)\ A * \mathsf{Exp}\ (E\ x)\ (t'' - t, \infty) \\ =\ &\textstyle\int_{(t', x')|t'' < t'} \mathsf{map}\ ((t', x'){\cdot})\ (N'\ (t', x'))A\ \mathsf{d}K\ (t, x) \end{aligned}$$

By combining both integrals and using the integral rule for the bind operator we proved Eq. (14). □

Transition probabilities We introduce *transition probabilities* $p\ x\ y\ t$, the probability that the Markov chain when started in x is at time t in state y. Due to the problem with explosion, we do not use trace. We introduce the inductively defined predicate $\mathsf{trace}_{\mathsf{in}}\ x\ t\ y\ \omega$ stating that when started in x, the trace ω is at time t in y.

$$\mathsf{trace}_{\mathsf{in}}\ x\ t\ y\ ((t', y'){\cdot}\omega) = \begin{cases} x = y & \textbf{if}\ t < t' \\ \mathsf{trace}_{\mathsf{in}}\ x\ t\ y'\ \omega & \textbf{otherwise} \end{cases}$$

$$p\ x\ y\ t = \mathsf{proc}_{\mathsf{stream}}\ (0, x)\ \{\omega \mid \mathsf{trace}_{\mathsf{in}}\ y\ t\ x\ \omega\}$$

For p to be well defined we need to show that $\mathsf{trace}_{\mathsf{in}}$ is measurable. This is done by representing it as a fixed point and showing that its functional is measurable.

If we interpret $p\ x\ y\ t$ as a t indexed matrix with column in x and row in y, then p forms a *matrix semi-group*, i.e. $p(t) * p(t') = p(t + t')$ for $0 \leq t, t'$. In Isabelle we state this more explicitly.

Theorem 7. *For $0 \leq t$ and $0 \leq t'$ we have*

$$p\ x\ y\ (t + t') = \sum_{x' \in I\ x} p\ x\ x'\ t * p\ x'\ y\ t'$$

Proof. The theorem is mostly proved using Theorem 6, however it requires some work to move the infinite sum through the occurring integrals. □

To analyse the transition probabilities we give the integration equation for p:

$$p\ x\ y\ t = \lfloor x = y \rfloor * e^{-t * E\ x} + \int_{u=0}^{t} E\ x * e^{(u-t)E\ x} \left(\int_{x'} p\ x'\ y\ u\ \mathsf{d}J\ x \right) \mathsf{d}\lambda \tag{15}$$

Where $\lfloor x = y \rfloor$ is 1 if $x = y$ otherwise it is 0. The integral $\int_{u=0}^{t} \ldots \mathsf{d}\lambda$ is the integral on the Lebesgue measure between 0 and t, i.e. the usual integral known from calculus. The integral $\int_{x'} \ldots \mathsf{d}J\ x$ integrates over the measure $J\ x$, i.e. over all values in $I\ x$ times the probability from $J\ x$.

The right-hand side of Eq. (15) has the following meaning: the first part of the sum is the probability that it stayed in x until time t, otherwise the integral integrates over the time u of the first jump into state x'. The proof is essentially done by unfolding one step of the jump & hold process and some calculus. This equation already implies that $p\ x\ y\ t$ is continuous in t (the Lebesgue integral is continuous in its boundary if the integrand is also integrable). From Eq. (15) follows that $p\ x\ y\ t$ is the solution to a differential equation.

Theorem 8 (Backward equation)**.** *Given $0 < t$ then the derivative $p'\ x\ y$ of $p\ x\ y$ in t fulfills the following equation:*

$$p'\ x\ y\ t = \left(\sum_{x' \in I\ x} R\ x\ x' * p\ x'\ y\ t \right) - E\ x * p\ x\ y\ t$$

The equation holds for the right-derivative of p if $t = 0$.

This proves not yet that a solution to the differential equation is unique. For finite state spaces this should be easy to derive from the work on flows of differential equations by (Immler and Traut 2016).

6. Conclusion and Discussion

We presented the formalization of discrete-time Markov processes and on top of them continuous-time Markov chains. The complete formalization has ~ 4300 lines, spread out over the Isabelle repository and the AFP entry (Hölzl and Nipkow 2012a). We used Markov kernels to describe the transition behaviour, and constructed the Markov kernel from the transition rates of a continuous-time Markov chain. The extension theorem by Ionescu-Tulcea gives us a straightforward way to show the existence of the Markov processes. Most proofs, especially Theorem 2, Theorem 4, and Theorem 6, involved a coinduction step or equality of measures reduced to generating sets, and then strongly relied on equational reasoning on probability measures.

Especially the Giry monad allows us to express most statements about Markov processes as equalities in the category of probability measures itself. This can be seen in the statements of Lemma 1, the construction of the extension theorem by Ionescu-Tulcea (Theorem 1), and the Theorems 2, 3, and 6. It is obviously possible either to give a more explicit construction of the involved measures or to express both sides of the statements as measures under concrete sets. But the usage of the Giry monad allows us to state these theorems and constructions in a concise way, and to manipulate them with the monad laws. Another part of the Giry monad is the introduction of measurable space on probability measures prob-algebra. With this Markov kernels are *just* measurable functions, hence the constructions and proof methods for measurable functions are immediately available.

Unfortunately, working with the Giry monad in Isabelle is not always straightforward. Many rewrite steps were performed manually as it required to instantiate specific measurable spaces only occurring in side conditions. An example is the associativity rule: $(M \gg\!= f) \gg\!= g = M \gg\!= (\lambda x.\ f\ x \gg\!=\ g)$. The rule has as side conditions that f and g are measurable: $f \in M \to_m \mathsf{prob\text{-}algebra}\ N$ and $g \in N \to_m \mathsf{prob\text{-}algebra}\ R$. The simplifier only accepts rewrite rules where all variables occur on the left-hand side of the equation, so N and R are problematic. An option would be to annotate bind with the result measurable space, as it is done for the return and map functions.[3] But this would render many statements very unreadable. A solution would be to represent measurable spaces as type classes (as done in (Berg 2013)). However this would force exactly one measurable space per type, making the formalization of product space, filtrations, and stopping times more complicated.

[3] These measurable spaces need to be written in Isabelle/HOL, but for clarity reasons were hidden in this paper.

Another problem in Isabelle/HOL's measure theory is the treatment of σ-algebras as measures. Only the type of measures is used to avoid a duplication when introducing product measures, the counting measure and the Lebesgue measure. However, it turns out that this complicates measurability proofs: instead of only applying typing-rules it needs to bring each measures into a normal form modulo sets.

A helpful tool is the *measurability prover*, a tool integrated into Isabelle's simplifier capable of proving many measurability statements. It employs some simple tools, for example removing terms of a countable type, i.e. $f\ (g\ x)\ x$ is reduced to measurability of $f\ n$ (for all n) and g when the range of g is countable, bringing measurability statements into a normal form, special handling for function application $f\ x\ i$ when $f\ x$ is in a product measurable space. With regard to measurablility, streams also have an advantage compared to sequences: functions on streams have one measurable space, while sequences are seen as product types. Measurability in product spaces adds one complication: we need to check membership of the indices, e.g. for the statement $(\lambda x.\ x_i) \in \Pi_{i\in I}\ M_i \to_m M_i$ we need to prove $i \in I$. This proof is not necessary on streams.

The construction mechanism for discrete-time Markov processes provides a powerful tool for constructing stochastic processes. With this mechanism many models can be built while avoiding invoking a complicated theorem like Caratheodory's extension theorem. For the future we want to build the current discrete-time Markov chains on top of Markov processes and extend the formalization of stopping times. With broad support for stopping times, we want to extend Theorem 6 to be a strong Markov property on continuous-time Markov chains. The proof should work similarly to the variant for discrete-time Markov processes. Also, to analyze continuous-time Markov chains with an infinite state space, we want to complement Theorem 8 with the proof that p is the minimal solution to the differential equation.

Acknowledgments

The author wants to thank the anonymous reviewers for their many insightful comments and suggestions. The author's work has been supported by the DFG projects Ni 491/15-1 and Ni 491/16-1.

References

P. Audebaud and C. Paulin-Mohring. Proofs of randomized algorithms in Coq. In *Special Issue on MPC 2006*, volume 74 of *Science of Computer Programming*, pages 568–589. 2009.

J. Avigad, J. Hölzl, and L. Serafin. A formally verified proof of the central limit theorem, 2016. Submitted to JAR in July 2016 (`https://arxiv.org/abs/1405.7012`).

M. Backes, M. Berg, and D. Unruh. A formal language for cryptographic pseudocode. In I. Cervesato, H. Veith, and A. Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Rea-*

soning (LPAR 2008), volume 5330 of *Lecture Notes in Computer Science*, pages 353–376. Springer, 2008. doi: 10.1007/978-3-540-89439-1_26.

C. Baier, B. R. Haverkort, H. Hermanns, and J.-P. Katoen. Model-checking algorithms for continuous-time markov chains. *IEEE Trans. Software Eng.*, 29(6):524–541, 2003. doi: 10.1109/TSE.2003.1205180.

M. Berg. *Formal Verification of Cryptographic Security Proofs*. PhD thesis, Saarland University, 2013.

J. C. Blanchette, J. Hölzl, A. Lochbihler, L. Panny, A. Popescu, and D. Traytel. Truly modular (co)datatypes for Isabelle/HOL. In G. Klein and R. Gamboa, editors, *Interactive Theorem Proving (ITP 2014)*, volume 8558 of *LNCS*, pages 93–110. Springer, 2014a.

J. C. Blanchette, A. Popescu, and D. Traytel. Unified classical logic completeness. In S. Demri, D. Kapur, and C. Weidenbach, editors, *Automated Reasoning (IJCAR 2014)*, volume 8562 of *LNCS*, pages 46–60. Springer, 2014b.

A. Clark, S. Gilmore, J. Hillston, and M. Tribastone. Stochastic process algebras. In M. Bernardo and J. Hillston, editors, *Formal Methods for Performance Evaluation*, pages 132–179. Springer, 2007. doi: 10.1007/978-3-540-72522-0_4.

D. Cock. Verifying probabilistic correctness in Isabelle with pGCL. In F. Cassez, R. Huuck, G. Klein, and B. Schlich, editors, *Systems Software Verification (SSV 2012)*, volume 102 of *EPTCS*, pages 167–178, 2012.

E.-E. Doberkat. *Stochastic relations: Foundations for Markov transition systems*. Studies in Informatics. Chapman & Hall/CRC, 2007.

M. Eberl, J. Hölzl, and T. Nipkow. A verified compiler for probability density functions. In *European Symposium on Programming (ESOP 2015)*, LNCS, 2015.

M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology and Analysis*, volume 915 of *Lecture Notes in Mathematics*, pages 68–85, 1982.

S. Gouezel. Ergodic theory. *The Archive of Formal Proofs*, Dec 2015. ISSN 2150-914x. `https://www.isa-afp.org/entries/Ergodic_Theory.shtml`, (Formal proof development).

T. C. Hales, M. Adams, G. Bauer, D. T. Dang, J. Harrison, T. L. Hoang, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, T. Q. Nguyen, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, A. H. T. Ta, T. N. Tran, D. T. Trieu, J. Urban, K. K. Vu, and R. Zumkeller. A formal proof of the kepler conjecture. *CoRR*, abs/1501.02155, 2015.

J. Hölzl. *Construction and Stochastic Applications of Measure Spaces in Higher-Order Logic*. PhD thesis, Technische Universität München, 2013.

J. Hölzl. Markov chains and Markov decision processes in Isabelle/HOL, 2016. Submitted to JAR in December 2015 (`http://in.tum.de/~hoelzl/mdptheory`).

J. Hölzl and A. Heller. Three chapters of measure theory in Isabelle/HOL. In M. C. J. D. van Eekelen, H. Geuvers, J. Schmaltz, and F. Wiedijk, editors, *Interactive Theorem Proving (ITP 2011)*, volume 6898 of *LNCS*, pages 135–151. Springer, 2011.

J. Hölzl and T. Nipkow. Markov models. *The Archive of Formal Proofs*, Jan 2012a. ISSN 2150-914x. `https://www.isa-afp.org/entries/Markov_Models.shtml` (Formal proof development).

J. Hölzl and T. Nipkow. Interactive verification of Markov chains: Two distributed protocol case studies. In *QFM 2012*, volume 103 of *EPTCS*, 2012b.

J. Hölzl and T. Nipkow. Verifying pCTL model checking. In C. Flanagan and B. König, editors, *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2012)*, volume 7214 of *LNCS*, pages 347–361, 2012c.

J. Hölzl, A. Lochbihler, and D. Traytel. A formalized hierarchy of probabilistic system types - proof pearl. In C. Urban and X. Zhang, editors, *Interactive Theorem Proving (ITP 2015)*, volume 9236 of *LNCS*, pages 203–220. Springer, 2015.

J. Hurd. *Formal Verification of Probabilistic Algorithms*. PhD thesis, 2002.

J. Hurd, A. McIver, and C. Morgan. Probabilistic guarded commands mechanized in HOL. *Theoretical Computer Science*, 346 (1):96–112, Nov 2005.

F. Immler. Generic construction of probability spaces for paths of stochastic processes in Isabelle/HOL. Master's thesis, Technical University of Munich, 2012. `http://home.in.tum.de/~immler/mastersthesis/index.html`.

F. Immler and C. Traut. The flow of ODEs. In C. J. Blanchette and S. Merz, editors, *Interactive Theorem Proving (ITP 2016)*, volume 9807 of *LNCS*, pages 184–199, 2016. doi: 10.1007/978-3-319-43144-4_12.

D. R. Lester. Topology in PVS: continuous mathematics with applications. In *Proceedings of the second workshop on Automated formal methods*, AFM '07, pages 11–20, 2007. doi: 10.1145/1345169.1345171.

L. Liu, O. Hasan, V. Aravantinos, and S. Tahar. Formal reasoning about classified Markov chains in HOL. volume 7998 of *LNCS*, pages 295–310. Springer, 2013.

J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilisitc Mathematics. Cambridge University Press, 1997.

D. Pollard. *A Users's Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2002.

K. S. Trivedi. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. Wiley, 2nd edition edition, 2002.