Analyzing Discrete-Time Markov Chains with Countable State Space in Isabelle/HOL

Johannes Hölzl*

Institut für Informatik, Technische Universität München
hoelzl@in.tum.de

Abstract. Discrete-time Markov chains are an important tool in probabilistic analysis of computer systems. For example they are used to describe the behavior of computer programs with probabilistic choice or the time-dependent distribution of input values. Current formalizations of Markov chains are restricted to a finite state space. We extend this to a countable state space, construct the stochastic process of a Markov chain given a matrix of transition probabilities, and prove the equivalence with the axiomatic definition as stochastic process. Based on this we introduce irreducible, recurrent, and aperiodic classes, generating functions and stationary distributions to analyze Markov chains.

Keywords: Markov chain, probability theory, mathematical analysis, Isabelle/HOL

1 Introduction

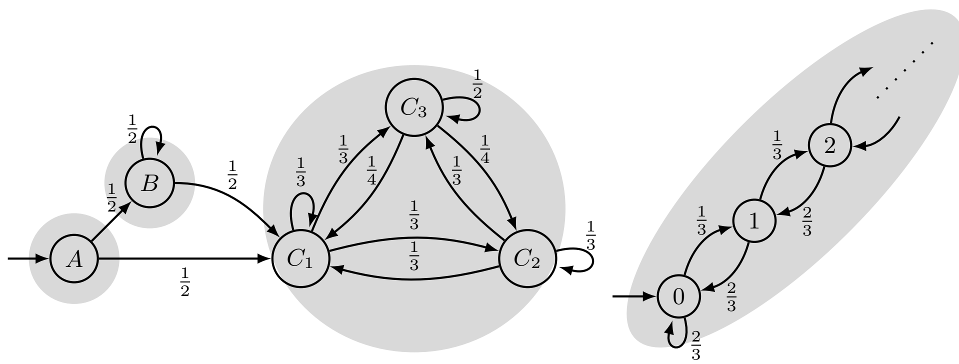
Modeling systems as discrete-time Markov chains is a popular technique to analyze probabilistic behavior of network protocols, algorithms, communication systems or biological systems. In computer science they are used to analyze probabilistic programs [4], or to analyze queuing and reliability problems [14].

A Markov chain describes the behavior of a probabilistic process. Similar to a state machine, the Markov chain has a set of states, transitions between these states and a starting state. The transitions are labelled with probabilities. The decision which state to choose next depends only on the current state, it is independent of previously visited states, and also independent of the time. In this paper we only handle Markov chains with discrete time (i.e. time values are natural numbers) and a discrete state space (i.e. the set of states is countable).

Fig. 1 shows the transition graphs of two Markov chains. The vertices are the states and the edges describe the non-zero transition probabilities. For (a) the starting state is A, here the process chooses in half of all cases B otherwise C_1. It stays in B with probability 1/2, otherwise it chooses C_1. In C_1 and in C_2 it chooses with probability 1/3 any C_i. It stays in C_3 with probability 1/2 and chooses C_2 or C_3 with probability 1/4.

* Supported by the DFG Projekt NI 491/15-1

Fig. 1 (b) shows a so called birth-death process with an infinite state space. With probability 1/3 it goes one state upwards, and with probability 2/3 it goes one state downwards (in the case of 0 it stays in 0 with probability 2/3).



(a) A finite process

(b) An infinite birth-death process

Fig. 1. Two Markov chains

These Markov chains already yields some questions:

Q1 When we start (b) does it always come back to 0?
Q2 Is it possible that (a) never reaches C_3 or that (b) never reaches a state n?
Q3 When we start (b) in 0, what is the average time until it reaches 0 again?
Q4 When (a) runs for a long time, is C_3 more likely to occur than C_1 or C_2?

To answer these questions we need to formalize probability and Markov chain theory. For readers new to this topic a introductory text book is [15], which we use for the analysis part of our formalization.

Our formalization [9] is split into three parts:

Measure and probability This is required as the set of all traces of a Markov chain is uncountable. This part provides us with construction mechanisms for measure spaces, like infinite products of probability spaces, and with the Lebesgue integral.

Defining Markov chains We provide two approaches: One assumes a state space, a matrix of transition probabilities, and a starting state. From this we constructs the probability space of traces. The other approach assumes the existence of a probability space and a time-indexed family of random variables describing the behavior of the Markov chain at a certain time. We show the equivalence between both approaches.

Analyzing Markov Chains Finally we use the trace space to give formal meaning to the questions **Q1** to **Q5**. We formalize concepts like irreducible classes, hitting times, and stationary distributions. We prove theorems to reduce quantitative questions to the analysis of the transition graph, and to the solution of a linear equation system.

2 Related Work

For the construction of Markov chains we Isabelle/HOL's measure and probability theory presented in [5]. This work constructs finite-state Markov chains to verify probabilistic model checking, and for reachability analysis of Markov chains defines in HOL. In this paper we extend this work to countable (possibly infinite) state spaces and further analysis concepts such as (positive) recurrence, the period of a state and stationary distributions. Popescu, Hölzl and Nipkow [13] use the work presented in this paper to model programs with probabilistic choice and parallel composition as infinite-state Markov chains.

One of the first formalizations of measure and probability theory in a HOL theorem prover was Hurd's probabilistic monad [7]. He formalized the foundations of measure theory and used it to construct a probability space on $\mathbb{N} \Rightarrow \mathbb{B}$. He verifies programs using a random number generator as functions reading bits from traces in this probability space.

Liu et al. [12] classify finite-state Markov chains. They axiomatically define Markov chains as stochastic processes. They show for aperiodic Markov chains that the limit of the marginal probability is a stationary distribution.

A simpler approach, which does not require measure theory, is to employ expectation transformers on countable distributions. Hurd, McIver and Morgan [8] use this approach, they even allow demonic (non-deterministic) choice. This work was recently ported to Isabelle/HOL by Cock [2]. Audebaud and Paulin-Mohring [1] develop a shallow embedding of probabilistic programs in Coq by formalizing a monad of discrete distributions.

A fully automatic approach to analyzing Markov chains is using probabilistic model checkers, like PRISM [11] or MRMC [10]. While model checkers work fully automatically they can only analyze problems on finite-state Markov chains, and their analysis is usually limited to reachability probabilities and average cost computation. Also, as they work non-symbolically, they suffer from state-space explosion.

3 Preliminaries

The term syntax used in this paper follows Isabelle/HOL, i.e. function application is juxtaposition as in $f\ t$ and $t :: \tau$ means that term t has type τ. Types are built from the base types $\mathbb{B}$ (booleans), $\mathbb{N}$ (natural numbers), $\mathbb{R}$ (reals), $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ (extended reals), type variables (α, β, etc), via the function type constructor $\alpha \Rightarrow \beta$, and via the set type constructor $\alpha\ \mathsf{set}$. We use $\Longrightarrow$ for logical implication; it binds – in contrast to Isabelle/HOL notation – stronger than universal quantification, i.e. $\forall x.\ P\ x \implies Q\ x$ equals $\forall x.\ (P\ x \implies Q\ x)$. We use two operations on traces $\omega :: \mathbb{N} \Rightarrow \alpha$: $\omega \circ_n \omega'$ to concatenate two traces:

if $i < n$ then $(\omega \circ_n \omega')\ i = \omega\ i$, otherwise $(\omega \circ_n \omega')\ (i + n) = \omega'\ i$, and $s \cdot \omega$ to prepend an element: $(s \cdot \omega)\ 0 = s$, and $(s \cdot \omega)\ (n + 1) = \omega\ n$. χ_A is the indicator function $\chi_A\ x = 1$ if $x \in A$ otherwise $\chi_A\ x = 0$. We abbreviate $A \to B = \{f \mid \forall a \in A.\ f \quad \in B\}$. $f\ z \xrightarrow{z \to 1^-} x$ states that the function $f\ z$ converges to x as z approaches 1 from the left. $X\ n \xrightarrow{n \to \infty} x$ states that the sequence $X\ n$ converges to x. We write $\mathit{LEAST}\ n.\ P\ n$ for the smallest natural number n fulfilling $P\ n$.

We use Isabelle's locale-mechanism [3] to define mathematical structures. The **locale**-command defines a new predicate L by giving a list of constants $c_1, c_2, \ldots, c_n$ and a list of assumptions $P_1, P_2, \ldots, P_k$:

locale $L =$
fixes $c_1 :: \alpha$ **and** $c_2 :: \beta$ **and** $\cdots$ **and** $c_n :: \gamma$
assumes P_1 **and** P_2 **and** $\cdots$ **and** P_k

This defines the predicate $L\ c_1\ \ldots\ c_n = P_1 \wedge P_2 \cdots P_k$. In Isabelle/HOL, this gives us a context where we have direct access to the assumptions $P_1, P_2, \ldots, P_k$ and to other theorems proved in this context.

4 Measure and Probability Theory

Before we start with Markov chains itself, we give a short overview of the required probability theory. We need a theory to (1) operate on measures (probabilities), (2) define integrals (expectations) on such measures, and (3) mechanisms to construct measures.

A *measure* consists of a measure space Ω, measurable sets $\mathcal{A}$ and a measure function μ. The measurable sets $\mathcal{A}$ form a σ-algebra consisting of subsets of Ω which are closed under countable union and complement. Measure functions are zero on the empty set $\mu\ \emptyset = 0$, non-negative $0 \leq \mu\ A$ for measurable sets $A \in \mathcal{A}$, and countably additive $\mu\ (\bigcup_i F_i) = \sum_i \mu\ F_i$, for a disjoint sequence F of measurable sets. We are mostly interested in probability measures where $\mu\ \Omega = 1$. In abstract math, we write $\Pr(P) = \mu\ \{\omega \in \Omega.\ P\ \omega\}$, i.e. the bound variable ω is not mentioned.

In Isabelle/HOL, measures are their own type α *measure*, with the projections

space	$:: \alpha$ *measure* $\Rightarrow \alpha$ *set*	for the measure space Ω,
sets	$:: \alpha$ *measure* $\Rightarrow \alpha$ *set set*	for the measurable sets $\mathcal{A}$, and
measure	$:: \alpha$ *measure* $\Rightarrow \alpha$ *set* $\Rightarrow \overline{\mathbb{R}}$	for the measure function μ.

If P is a *probability space* we write *prob-space* $P \iff$ *measure* P (*space* P) $= 1$. We write $\Pr(\omega$ *in* $M.\ P\ \omega) =$ *measure*$_{\mathbb{R}}$ $M\ \{\omega \in$ *space* $M.\ P\ \omega\}$, where *measure*$_{\mathbb{R}}$ is the restriction of *measure* to $\mathbb{R}$. We also define the conditional probability $\Pr(\omega$ *in* $M.\ P\ \omega \mid Q\ \omega) = \Pr(\omega$ *in* $M.\ P\ \omega \wedge Q\ \omega) / \Pr(\omega$ *in* $M.\ Q\ \omega)$. If $\Pr(\omega$ *in* $M.\ Q\ \omega) = 0$ then $\Pr(\omega$ *in* $M.\ P\ \omega \mid Q\ \omega) = 0$.

For a function $f \in$ *measurable* $M\ N$ all sets $\{\omega \in$ *space* $M.\ f\ \omega \in A\}$ are measurable in M for all $A \in$ *sets* N. We write $P \in$ *measurable* $M\ \mathbb{B}$, when

$\{\omega \in \mathsf{space}\ M.\ P\ \omega\}$ is measurable, and $f \in \mathsf{measurable}\ M\ \mathsf{borel}$, when $\{\omega \in \mathsf{space}\ M.\ f\ \omega < a\}$ is measurable for all $a :: \mathbb{R}$ or $a :: \overline{\mathbb{R}}$. In probability theory measurable functions are also called random variables.

For the Lebesgue integral we have two functions: $\int_x^+ f\ x\ dM :: \overline{\mathbb{R}}$ for non-negative functions $f :: \alpha \Rightarrow \overline{\mathbb{R}}$ which might be infinite, and $\int_x f\ x\ dM :: \mathbb{R}$ for real-valued functions $f :: \alpha \Rightarrow \mathbb{R}$ which is only defined if the integral is real. We have the usual rules: linearity $\int_x^+ c \cdot f\ x + d \cdot g\ x\ dM = c \cdot (\int_x^+ f\ x\ dM) + d \cdot (\int_x^+ g\ x\ dM)$ and monotone convergence $\int_x^+ f_n\ x\ dM \xrightarrow{n\to\infty} \int_x^+ \sup_n f_n\ x\ dM$. Similar to measures the integrals need measurable functions, the previous rules only hold for functions $f, g \in \mathsf{measurable}\ M\ \mathsf{borel}$, and for an increasing sequence of Borel-measurable functions f_n. In probability theory the integral on a probability space M is also called expectation.

4.1 Constructing Measure Spaces

Constructing measure spaces can be unwieldy, for this reason we provide a couple of construction mechanisms for measure spaces. In this section we assume measurability of the involved sets and functions. See [5] for a detailed description.

The counting measure $\mathsf{count}\ S$ for a countable set S assigns each set $A \subseteq S$ its cardinality as measure. The integral is equal to the sum of the integrand over S. This integral equation only works when the support of f is finite. In Isabelle/HOL there is no sum operation over arbitrary countable sets, instead we take advantage of the integral over $\mathsf{count}\ S$.

$$\begin{array}{lll}
\mathsf{count} & :: \alpha\ \mathsf{set} \Rightarrow \alpha\ \mathsf{measure} & \\
\mathsf{measure}\ (\mathsf{count}\ S)\ A & = \mathsf{if\ finite}\ A\ \mathsf{then\ card}\ A\ \mathsf{else}\ \infty & \\
\displaystyle\int_x^+ f\ x\ d\mathsf{count}\ S & = \displaystyle\sum_{x \in S \mid f\ x \neq 0} f\ x & \text{if } \{x \in S \mid f\ x \neq 0\} \text{ is finite}
\end{array}$$

A function $X \in \mathsf{measurable}\ M\ N$ induces a measure on N, the so called *push-forward measure* or *distribution* of X.

$$\begin{array}{ll}
\mathsf{distr} & :: \alpha\ \mathsf{set} \Rightarrow \beta\ \mathsf{set} \Rightarrow (\alpha \Rightarrow \beta) \Rightarrow \beta\ \mathsf{measure} \\
\mathsf{measure}\ (\mathsf{distr}\ M\ N\ X)\ A & = \mathsf{measure}\ M\ \{\omega \in \mathsf{space}\ M \mid X\ \omega \in A\} \\
\displaystyle\int_x^+ g\ x\ d\mathsf{distr}\ M\ N\ X & = \displaystyle\int_x^+ g\ (X\ x)\ dM
\end{array}$$

We can use non-negative, Borel-measurable functions f as *densities*, i.e. associate to each element of the measure space a density (or weight). We use the Lebesgue integral to compute the weighted measure of a set.

$$\begin{array}{ll}
\mathsf{density} & :: \alpha\ \mathsf{set} \Rightarrow (\alpha \Rightarrow \overline{\mathbb{R}}) \Rightarrow \alpha\ \mathsf{measure} \\
\mathsf{measure}\ (\mathsf{density}\ M\ f)\ A & = \displaystyle\int_x^+ f\ x \cdot \chi_A\ x\ dM \\
\displaystyle\int_x^+ g\ x\ d\mathsf{density}\ M\ f & = \displaystyle\int_x^+ g\ x \cdot f\ x\ dM
\end{array}$$

We use the density measure in combination with the counting space to construct the *point measure* of p on the space S.

$$\begin{aligned}
&\mathit{point} && :: \alpha\ \mathit{set} \Rightarrow (\alpha \Rightarrow \overline{\mathbb{R}}) \Rightarrow \alpha\ \mathit{measure} \\
&\mathit{point}\ S\ p && = \mathit{density}\ (\mathit{count}\ S)\ p \\
&\mathit{measure}\ (\mathit{point}\ S\ p)\ A && = \textstyle\sum_{x \in A} p\ x \qquad \text{if } A \text{ is finite}
\end{aligned}$$

In probability theory one often needs infinitely many independent random variables. We construct them using the product of infinitly many probability measures. While the index set I is allowed to be uncountable, the set J in the following equation needs to be finite.

$$\Pi_{:} : \iota\ \mathit{set} \Rightarrow (\iota \Rightarrow \alpha\ \mathit{measure}) \Rightarrow (\iota \Rightarrow \alpha)\ \mathit{measure}$$
$$\Pr(\omega\ \mathit{in}\ (\Pi_{i \in I} M\ i)\,.\ \forall i \in J.\ \omega\ i \in A\ i) = \textstyle\prod_{i \in J} \Pr(x\ \mathit{in}\ M\ i.\ x \in A\ i)$$

The construction of this measure space is complicated, it requires Caratheodory's extension theorem. It will be very helpful in the next section.

5 The Markov Chain Trace Space

To answer the questions **Q1** to **Q4** our informal description given in the introduction is not enough. For a formal approach we need a probability space assigning probabilities to sets of traces in the Markov chain.

5.1 Axiomatic Definition as Stochastic Processes

One way to represent Markov chains is as a *stochastic process* $X \in \mathbb{N} \to \Omega \to S$, i.e. a family of functions $(X_t)_{t \in \mathbb{N}}$ where X_t is measurable on a probability space M. The space M represents the probabilistic behavior of our entire "world" and with X we observe the behavior of our Markov chain.

We introduce the locale *Discrete-Time-Markov-Chain*. It assumes a stochastic process X and a probability space M which has the properties of a Markov chain, i.e. it is memoryless and time-homogeneous.

locale *Discrete-Time-Markov-Chain* =
  fixes $M :: \alpha\ \mathit{measure}$ **and** $S :: \beta\ \mathit{set}$ **and** $X :: \mathbb{N} \Rightarrow \alpha \Rightarrow \beta$ +
  assumes *prob-space* M **and** *countable* S
  assumes $\forall t.\ X\ t \in \mathit{measurable}\ M\ (\mathit{count}\ S)$

  assumes $\forall t\ x\ y.$ — The stochastic process X is *memoryless*:
    $\Pr(\omega\ \mathit{in}\ M.\ \forall t' \le t.\ X\ t'\ \omega = y\ t') \ne 0 \implies$
    $\Pr(\omega\ \mathit{in}\ M.\ X\ (t+1)\ \omega = x \mid \forall t' \le t.\ X\ t'\ \omega = y\ t') =$
    $\Pr(\omega\ \mathit{in}\ M.\ X\ (t+1)\ \omega = x \mid X\ t\ \omega = y\ t)$

  assumes $\forall t\ t'\ x\ y.$ — The stochastic process X is *time-homogeneous*:
    $\Pr(\omega\ \mathit{in}\ M.\ X\ t\ \omega = y) \ne 0\ \wedge\ \Pr(\omega\ \mathit{in}\ M.\ X\ t'\ \omega = y) \ne 0 \implies$
    $\Pr(\omega\ \mathit{in}\ M.\ X\ (t+1)\ \omega = x \mid X\ t\ \omega = y) =$
    $\Pr(\omega\ \mathit{in}\ M.\ X\ (t'+1)\ \omega = x \mid X\ t'\ \omega = y)$

The assumptions that the conditions of the conditional probabilities are non-zero are required. Otherwise the Markov chain in Fig. 1 (a) would not fulfill our locale: as $\Pr(X_t = A) = 0$ for all $0 < t$, an unrestricted time-homogeneous property would imply $\frac{1}{2} = \Pr(X_1 = B \mid X_0 = A) = \Pr(X_2 = B \mid X_1 = A) = 0$. It is not necessary to restrict x and y to be elements of S, as the probability to reach elements outside S is zero.

5.2 Constructive Definition with Transition Probabilities

Often a Markov chain is not given as a stochastic process, but as a state space S, the initial distribution $\iota\ x = \Pr(X_0 = x)$, and the transition matrix $\tau\ x\ y = \Pr(X_{t+1} = x \mid X_t = y)$ (this equation holds only for $\Pr(X_t = y) > 0$). With the mechanisms from Section 4.1 we now construct a probability space and a stochastic process having a specific transition matrix and initial distribution.

We assume a countable state space $S :: \alpha\,\mathit{set}$ and a matrix of transition probabilities. The matrix is not formalized as a function $\alpha \Rightarrow \alpha \Rightarrow \mathbb{R}$, but as a so called Markov kernel $K :: \alpha \Rightarrow \alpha\,\mathit{measure}$. This exploits the fact, that each row in τ is a probability distribution. We assume no initial distribution, however when measuring traces we provide a starting state.

locale *Discrete-Markov-Kernel* =
  fixes $S :: \alpha\,\mathit{set}$ **and** $K :: \alpha \Rightarrow \alpha\,\mathit{measure}$
  assumes *countable* S **and** $S \neq \emptyset$
  assumes $\forall x.\ \mathit{sets}\ (K\ x) = \mathcal{P}(S) \wedge \mathit{prob\text{-}space}\ (K\ x)$

For the rest of this section we assume a discrete Markov kernel with state space S and kernel K. The transition probability $\tau\ x\ y$ to go from state x to y is written as $\mathit{measure}\ (K\ x)\ \{y\}$.

Our goal is now to define a probability measure $\mathcal{P}_x$ on the space of all traces $\mathbb{N} \to S$, where x is the starting state. The trace space $\mathcal{P}_x$ should assign the probability $\tau\ x\ \omega_0 \cdot \tau\ \omega_0\ \omega_1 \cdots \tau\ \omega_{n-1}\ \omega_n$ to the set of all traces starting with $\omega_0, \omega_1, \ldots, \omega_n$.

How can we proof the existence of such a measure? Our first option is to use the method by Hurd [7] and define the algebra of finite unions of cylinders, i.e. sets of traces starting with the same prefix. We proved Caratheodory's extension theorem, but to operate on finite unions of cylinders it is very cumbersome.

An alternative is to cast the probability measure $\mathcal{P}_x$ out of the infinite product measure with $\mathbb{N}$ as index set and a probability space on $S \to S$ as factors. The nested product $\mathit{\Pi}_{i\in\mathbb{N}}(\mathit{\Pi}_{y\in S} K\ y) :: (\mathbb{N} \Rightarrow \alpha \Rightarrow \alpha)\,\mathit{measure}$ is indexed by time n and then by the current state y. We simply map an element from this product space into a trace in $\mathcal{P}_x$, by starting with x and then following the states selected at each time point. For this we define the function *trace*:

$$\begin{aligned}
&\mathit{trace} && :: \alpha \Rightarrow (\mathbb{N} \Rightarrow \alpha \Rightarrow \alpha) \Rightarrow (\mathbb{N} \Rightarrow \alpha)\\
&\mathit{trace}\ x\ \pi\ 0 && = \pi\ 0\ (\mathbf{if}\ x \in S\ \mathbf{then}\ x\ \mathbf{else}\ \mathbf{SOME}\ x.\ x \in S)\\
&\mathit{trace}\ x\ \pi\ (n+1) && = \pi\ (n+1)\ (\mathit{trace}\ x\ \pi\ n)
\end{aligned}$$

By induction we have $\mathsf{trace}\, x \in \mathsf{measurable}\, (\Pi_{i\in\mathbb{N}} \Pi_{y\in S} K\ y)\ S_\Pi$. Here S_Π is the σ-algebra of infinite sequences of S: $S_\Pi = \Pi_{i\in\mathbb{N}} \mathsf{count}\, S$.

Definition 1 (Trace space). *As the function* $\mathsf{trace}\ x$ *is measurable, we use* $\mathsf{trace}\, x$ *to define the measure* $\mathcal{P}_x$*:*

$$\begin{aligned}
&\mathcal{P}\ \ :: \alpha \Rightarrow (\mathbb{N} \Rightarrow \alpha)\ \mathsf{measure}\\
&\mathcal{P}_x = \mathsf{distr}\, (\Pi_{i\in\mathbb{N}} \Pi_{y\in S} K\ y)\ S_\Pi\ (\mathsf{trace}\, x)
\end{aligned}$$

With this the measurable sets of $\mathcal{P}_x$ are the measurable sets of S_Π, i.e. the smallest σ-algebra where $\{\omega \in \mathbb{N} \to S \mid \omega\ i \in A\}$ is measurable for all $A \subseteq S$ and all $i \in \mathbb{N}$. This trace space also contains traces ω with $\tau\ (\omega\ i)\ (\omega\ (i+1)) = 0$. In textbooks this is often avoided, but it simplifies proving measurability.

From the construction of $\mathcal{P}_x$ we derive equations to split expectations and probabilities of traces at an arbitrary time point n. These equations are similar to the Chapman-Kolmogorov equations, but more general:

Lemma 2 (Splitting rules). *We assume that* x *is in* S *and that* f *and* P *are measurable, then*

$$\begin{aligned}
\int^+_\omega f\ \omega\ d\mathcal{P}_x \qquad &= \int^+_\omega \left(\int^+_{\omega'} f\ (\omega \circ_n \omega')\ d\mathcal{P}_{(x\cdot\omega)\ n} \right) d\mathcal{P}_x \ \textit{and}\\
\Pr(\omega\ \mathsf{in}\ \mathcal{P}_x.\ P\ \omega) &= \int^+_\omega \Pr\left(\omega'\ \mathsf{in}\ \mathcal{P}_{(x\cdot\omega)\ n}.\ P\ (\omega \circ_n \omega')\right) d\mathcal{P}_x.
\end{aligned}$$

From this we can deduce a generic form of Chapman-Kolmogorov:

Lemma 3 (Chapman-Kolmogorov). *We assume that* x *is in* S *and that* P *and* Q_1 *are measurable. Lets assume that* P *can be split into the predicates* Q_1 *and* Q_2*:* $\omega, \omega' \in \mathsf{space}\, S_\Pi \implies P\ (\omega \circ_n \omega') \Longleftrightarrow Q_1\ \omega \wedge Q_2\ (((x \cdot \omega)\ n) \cdot \omega')$*, then*

$$\begin{aligned}
&\Pr(\omega\ \mathsf{in}\ \mathcal{P}_x.\ P\ \omega) =\\
&\quad \int_y \Pr\left(\omega\ \mathsf{in}\ \mathcal{P}_x.\ Q_1\ \omega \wedge (x \cdot \omega)\ n = y\right) \cdot \Pr\left(\omega\ \mathsf{in}\ \mathcal{P}_y.\ Q_2\ (t \cdot \omega)\right) d\mathsf{count}\, S.
\end{aligned}$$

5.3 Equivalence

We show that both definitions of a Markov chain are equivalent. With the splitting rules we show the memoryless and the time-homogeneous property.

Theorem 4 (The trace space $\mathcal{P}_x$ is a Markov chain).
If $\mathsf{Discrete\text{-}Markov\text{-}Kernel}\ S\ K$*, then the trace space* $\mathcal{P}_x$ *fulfills the Markov chain properties:* $\mathsf{Discrete\text{-}Time\text{-}Markov\text{-}Chain}\ \mathcal{P}_x\ S\ (\lambda n\ \omega.\ \omega\ n)$ *for* $x \in S$*.*

If we have a stochastic process X fulfilling the Markov chain properties, it is more complicated. The stochastic process may have an arbitrary initial distribution $\Pr(X_0 = s)$, whereas *Discrete-Markov-Kernel* assumes a single starting state. We construct a Markov kernel by embedding the state space S in the option type and use *None* as new starting state.

Definition 5 (Markov kernel from the stochastic process X). *We define $S' :: \beta$ option set to be an extension of $S :: \beta$ set.*

$$\begin{aligned}
&S' && :: \beta\ \textsf{option set}\\
&S' && = \{\textsf{None}\} \cup \{\textsf{Some}\ x \mid x \in S\}\\
&K' && :: \beta\ \textsf{option} \Rightarrow \beta\ \textsf{option measure}\\
&\textsf{measure}\ (K'\ \textsf{None})\ A && = \Pr(\omega\ \textsf{in}\ M.\ \textsf{Some}\ (X\ 0\ \omega) \in A)\\
&\textsf{measure}\ (K'\ (\textsf{Some}\ x))\ A && = \Pr(\omega\ \textsf{in}\ M.\ \textsf{Some}\ (X\ (t+1)\ \omega) \in A | X\ t\ \omega = x)
\end{aligned}$$

The last equation requires that $\Pr(\omega\ \textsf{in}\ M.\ X\ t\ \omega = x) \neq 0$. Together with the time-homogeneous property K' is well-defined.

With this definition we prove *Discrete-Markov-Kernel* $S'\ K'$. The resulting trace space $\mathcal{P}_{None}$ is equal to the distribution of X lifted to S'_{Π}:

Theorem 6 (The Markov kernel K' has the distribution of X).

$$\mathcal{P}_{None} = \textsf{distr}\ M\ S'_{\Pi}\ (\lambda\omega\ n.\ \textsf{Some}\ (X\ n\ \omega))$$

Theorem 4 and 6 show that for each Markov chain defined as a stochastic process exists a Markov kernel, and vice versa.

6 Analyzing Markov Chains

With $\mathcal{P}_x$ (for probabilities) and with the Lebesgue integral (for expectation) we formally restate the questions from the introduction:

Q1 When we start (b) does it always come back to 0?
$\Pr(\omega\ \textsf{in}\ \mathcal{P}_0.\ \exists t.\ \omega\ t = 0) = 1$?

Q2 Is it possible that (a) never reaches C_3 or that (b) never reaches a state n?
$\Pr(\omega\ \textsf{in}\ \mathcal{P}_A.\ \forall t.\ \omega\ t \neq C_3) \neq 0$ or $\Pr(\omega\ \textsf{in}\ \mathcal{P}_0.\ \forall t.\ \omega\ t \neq n) \neq 0$.

Q3 When we start (b) in 0, what is the average time until it reaches 0 again?
$\int^+_\omega \inf\ \{t \mid \omega\ t = 0\} d\mathcal{P}_0 =$?

Q4 When (a) runs for a long time, is C_3 more likely to occur than C_1 or C_2?
$\lim_{t\to\infty} \Pr(\omega\ \textsf{in}\ \mathcal{P}_0.\ \omega\ t = 0) > \lim_{t\to\infty} \Pr(\omega\ \textsf{in}\ \mathcal{P}_0.\ \omega\ t = n)$

Note that **Q1** is similar to **Q2**: **Q1** is the negation of **Q2**, but in **Q1** the starting state and the stopping state are equal.

Fortunately, Markov chain theory allows us to reduce these questions to an analysis of the graph of non-zero transitions, and to the solution of a linear equation system. In this section we formalize the required Markov chain theory. All following definitions and lemmas are in the locale *Discrete-Markov-Kernel* $S\ K$.

6.1 Accessibility

The non-zero transitions of a Markov chain define a directed graph. Properties of this directed graph are called *qualitative properties* of the Markov chain. For a state x, we define the set of *enabled*, *accessible* and *communicating* states and show their relation with probabilities on the trace space.

Definition 7 (Enabled states). *We define $E\ x$ as the set of all states reachable in one step with a non-zero transition probability:*

$$E :: \alpha \Rightarrow \alpha\ \mathsf{set}$$
$$y \in E\ x \iff y \in S \wedge \mathsf{measure}\ (K\ x)\ \{y\} \neq 0$$

Not every trace has everywhere enabled states, but *almost every* trace has:

Theorem 8. *Almost every trace is everywhere enabled*

$$x \in S \implies \Pr(\omega\ \mathsf{in}\ \mathcal{P}_x.\ \forall i.\ \omega\ i \in E\ ((x \cdot \omega)\ i)) = 1$$

Definition 9 (Accessible states). *The states accessible from x are inductivly defined as reflexive and transitive closure of E:*

$$\mathsf{accessible} :: (\alpha \times \alpha)\ \mathsf{set}$$
$$x \in S \implies (x, x) \in \mathsf{accessible}$$
$$(x, y) \in \mathsf{accessible} \wedge z \in E\ y \implies (x, z) \in \mathsf{accessible}$$

A state x is *accessible* from a state y iff $\Pr(y \text{ is reached}) > 0$:

Theorem 10. *Accessibility as probability on traces*

$$(x, y) \in \mathsf{accessible} \iff \Big(x \in S \wedge y \in S \wedge \exists n.\ \Pr(\omega\ \mathsf{in}\ \mathcal{P}_x.\ \omega\ n = y) \neq 0\Big)$$

Definition 11 (Communicating states). *The communicating relation is the symmetric variant of the accessibility relation:*

$$\mathsf{communicating} :: (\alpha \times \alpha)\ \mathsf{set}$$
$$(x, y) \in \mathsf{communicating} \iff (x, y) \in \mathsf{accessible} \wedge (y, x) \in \mathsf{accessible}$$

The relation *communicating* is an equivalence relation as it is symmetric, reflexive, and transitive. Its equivalence classes are called *irreducible*. We write the set of irreducible classes as $S/\mathsf{communicating}$. In Fig. 1 (a) the irreducible classes are $\{A\}$, $\{B\}$, and $\{C_1, C_2, C_3\}$. Fig. 1 (b) has only one irreducible class: $\{0, 1, 2, \ldots\}$. It is a so-called irreducible Markov chain.

An equivalence class is *essential* iff no state outside of the class is accessible by a state inside the class. If the Markov chain reaches a state in an essential class, it will stay in this class forever. Essential classes are equivalent to *strongly connected components* in graph theory.

Definition 12 (Essential class). *An* essential class C *is an irreducible class where all accessible states from C are again in C:*

$$\begin{aligned} \mathsf{essential} &:: \alpha\ \mathsf{set} \Rightarrow \mathbb{B} \\ \mathsf{essential}\ C &\iff C \in S/\mathsf{communicating}\ \wedge \\ &\qquad \forall (x, y) \in accessible.\ x \in C \implies y \in C \end{aligned}$$

6.2 Recurrence

Many interesting properties on Markov chains are related to *recurrence*, i.e. the probability to reach a state when started in this state.

Definition 13 (Recurrent states). *A state x is* recurrent *iff it is guaranteed that when we start in x, that we come back in x.*

$$\textit{recurrent}\ x \Longleftrightarrow \Pr(\omega\ \textit{in}\ \mathcal{P}_x.\ \exists n.\ \omega\ n = x) = 1$$

Q1 can be written as *recurrent* 0. For **Q3** we are interested in the average time to return to 0. For this we define the *hitting time* of x on the trace ω (this is also called *first passage time*):

Definition 14 (Hitting time). *The hitting time $t\ x\ \omega$ is the first occurence of x plus 1 if x is in ω. If x never occurs the hitting time is ∞.*

$$\begin{aligned} &t && :: \alpha \Rightarrow (\mathbb{N} \Rightarrow \alpha) \Rightarrow \overline{\mathbb{R}} \\ &t\ x\ \omega = && \textbf{\textit{if}}\ \exists n.\ \omega\ n = x\ \textbf{\textit{then}}\ (\textit{LEAST}\ n.\ \omega\ n = x) + 1\ \textbf{\textit{else}}\ \infty \end{aligned}$$

(Note that the hitting time is never 0)

Definition 15 (Average hitting time).

$$\begin{aligned} &M && :: \alpha \Rightarrow \overline{\mathbb{R}} \\ &M\ x = && \int_\omega^+ t\ x\ \omega\ d\mathcal{P}_x \end{aligned}$$

With this **Q3** can also be written as $M\ x$. When x is non-recurrent, $M\ x$ is always infinite. But even for a recurrent state it may be infinite. If $M\ x$ is real, the state is called positive recurrent:

Definition 16 (Positive recurrent states).

$$\begin{aligned} &\textit{pos-recurrent} && :: \alpha \Rightarrow \mathbb{B} \\ &\textit{pos-recurrent}\ x && \Longleftrightarrow \textit{recurrent}\ x \wedge M\ x < \infty \end{aligned}$$

If a recurrent state is not positive recurrent it is called a *null recurrent* state.

6.3 Quantities and Generating Functions

In the previous section we gave the definition of recurrence. Now we provide tools to relate (positive) recurrent states with other quantities, like the probability to hit a state infinitely often or the average number of reaching a state.

Definition 17 (Unbound quantities).

$$\begin{aligned} &G && :: \alpha \Rightarrow \alpha \Rightarrow \overline{\mathbb{R}} \qquad \textit{and}\ \ H, U :: \alpha \to \alpha \to \mathbb{R} \\ &G\ x\ y = && \int_\omega^+ \textit{measure}\ (\textit{count}\ \mathbb{N})\ \{i \mid (x \cdot \omega)\ i = y\}\ d\mathcal{P}_x \\ &H\ x\ y = && \Pr(\omega\ \textit{in}\ \mathcal{P}_x.\ \textit{infinite}\ \{i \mid \omega\ i = y\}) \\ &U\ x\ y = && \Pr(\omega\ \textit{in}\ \mathcal{P}_x.\ \exists n.\ \omega\ n = y) \end{aligned}$$

Obviously, $\textit{recurrent}\ x \Longleftrightarrow U\ x\ x = 1$.

In the definition of G, *measure* (*count* $\mathbb{N}$) A is used as cardinality of the set A which is ∞ if A is infinite. The usual cardinality in Isabelle/HOL is on natural numbers, hence it is 0 if A is infinite. With the exception of H the unbound quantities are expressible as infinite sums over time-bounded probabilities:

Definition 18 (Time-bounded reachability probabilities).

$$
\begin{aligned}
&p, u \quad :: \alpha \Rightarrow \alpha \Rightarrow \mathbb{N} \Rightarrow \mathbb{R} \\
&p\ x\ y\ n = \Pr(\omega\ \textit{in}\ \mathcal{P}_x.\ (x \cdot \omega)\ n = y) \\
&u\ x\ y\ n = \Pr(\omega\ \textit{in}\ \mathcal{P}_x.\ (\forall i < n.\ \omega\ i \neq y) \wedge \omega\ n = y)
\end{aligned}
$$

Lemma 19 (Unbound quantities as series of time-bounded probabilities).

$$G\ x\ y = \sum_n p\ x\ y\ n \quad M\ x = \sum_n u\ x\ x\ n \cdot (n+1) \quad U\ x\ y = \sum_n u\ x\ y\ n$$

The expectation $G\ x\ y$ can be infinite. To avoid case distinctions for proofs about G, and to take advantage of mathematical analysis, Markov chain theory introduces *generating function.* For a probability or expectation $P = \sum_n f_n$, its power series is called a generating function $P_g\ z = \sum_n f_n \cdot z^n$. Here z is a real number between 0 and 1. If z approaches 1 from the left, we have $P_g\ z \xrightarrow{z \to 1^-} P$.

Definition 20 (Generating functions for G, M, U, and F).

$$
\begin{aligned}
&G_g, M_g, U_g :: \alpha \Rightarrow \alpha \Rightarrow \mathbb{R} \Rightarrow \mathbb{R} \\
&G_g\ x\ y\ z = \sum_n p\ x\ y\ n \cdot z^n \qquad M_g\ x\ z = \sum_n u\ x\ x\ n \cdot (n+1) \cdot z^n \\
&U_g\ x\ y\ z = \sum_n u\ x\ y\ n \cdot z^{n+1}
\end{aligned}
$$

These functions are real valued. As long as $|z| < 1$, the sums are well-defined.

Lemma 21 (Unbound quantities as limits of generating functions). *The generating functions tend to their probabilistic equivalence, if z approaches 1 from the left:*

$$G_g\ x\ y\ z \xrightarrow{z \to 1^-} G\ x\ y \qquad M_g\ x\ z \xrightarrow{z \to 1^-} M\ x \qquad U_g\ x\ y\ z \xrightarrow{z \to 1^-} U\ x\ y$$

(Both rules for $M\ x$ require that x is recurrent.)

For example, we can show that the average number of visits to x is inversely proportional to the probability to *not* reach x:

$$G\ x\ x = 1/(1 - U\ x\ x) \quad .$$

But this equation is not true when $U\ x\ x = 1$. With generating functions we have a nice tool to relate them even in this case.

Theorem 22 (Relating *recurrent* and G). *Assume $x \in S$, then:*

$$G_g\ x\ x\ z = \frac{1}{1 - U_g\ x\ x\ z} \qquad \textit{for } |z| < 1$$

As $G_g\ x\ x\ z \xrightarrow{z \to 1^-} G\ x\ x$ *we also know:*

$$\frac{1}{1 - U_g\ x\ x\ z} \xrightarrow{z \to 1^-} G\ x\ x$$

So, $\mathsf{recurrent}\ x$ *(i.e.* $U\ x\ x = 1$*) is equal to* $G\ x\ x = \infty$*.*

As $G\ x\ x = \infty \iff G\ y\ y$ if $(x, y) \in \mathsf{communicating}$, we have:

Corollary 23 (*recurrent* is invariant on irreducible classes).

$$(x, y) \in \mathsf{communicating} \implies \mathsf{recurrent}\ x \iff \mathsf{recurrent}\ y$$

Recurrence is not only invariant on irreducible classes, we also now that recurrence of a class implies that it is an essential class. For this we relate recurrence with H and U:

Lemma 24 (Relation between *recurrent* and H and U).
Assume $x, y \in S$*, then:*

$$H\ x\ x = \textbf{if}\ \mathsf{recurrent}\ x\ \textbf{then}\ 1\ \textbf{else}\ 0$$
$$H\ x\ y = U\ x\ y \cdot H\ y\ y$$

Assume $\mathsf{recurrent}\ x$ *and* $(x, y) \in \mathsf{accessible}$*, then*

$$U\ y\ x = 1, \qquad \mathsf{recurrent}\ y, \quad \text{and} \qquad (x, y) \in \mathsf{communicating}$$

For recurrent states accessible equals communicating hence they form an essential class:

Theorem 25 (Recurrent classes are essential).

$$\mathsf{recurrent}\ x \implies \mathsf{essential}\ \{y \mid (x, y) \in \mathsf{accessible}\}$$

With Theorem 25 and the pigeon hole principle follows also that finite essential classes are recurrent:

Corollary 26 (Finite essential classes are recurrent).

$$\mathsf{finite}\ C \wedge C \in S/\mathsf{communicating} \implies \mathsf{essential}\ C \iff (\forall x \in C.\ \mathsf{recurrent}\ x)$$

Lemma 24 also helps us to answer **Q2** (which is $1-U\ A\ C_3 \neq 0$ or $1-U\ 0\ n \neq 0$), if we can show that C_3 or n are recurrent.

Similar to recurrence, positive recurrence is invariant on irreducible classes. To prove this, we show that the derivative of the generating function U_g is M_g and then applying l'Hôpital's rule to relate M and U:

Lemma 27 (Relation between M and U).

$$\mathsf{DERIV}\ (U_g\ x\ x)\ z\ :> M_g\ x\ z \qquad \text{for } |z| < 1 \text{ and}$$

$$\frac{1 - z}{1 - U_g\ x\ x\ z} \xrightarrow{z \to 1^-} \frac{1}{M\ x}\quad .$$

Hence positive recurrence is invariant on irreducible classes: either all states are positive recurrent or all states are null recurrent.

Corollary 28 (Positive recurrent is invariant on irreducible classes).

$$(x, y) \in \mathsf{communicating} \implies \mathsf{pos\text{-}recurrent}\ x \iff \mathsf{pos\text{-}recurrent}\ y$$

6.4 Stationary Distribution

A stationary distribution μ is a measure invariant under multiplication with τ: $\forall y \in S.\ \mu\ y = (\sum_{x \in S} \mu\ x \cdot \tau\ x\ y)$. In Isabelle/HOL, we use the integral as sum, and instead of the quantifier we use equality on measures[1]:

Definition 29 (Stationary distribution).

$$\begin{array}{l} \textsf{stationary-distribution} \quad :: \quad \alpha\ \textsf{measure} \Rightarrow \mathbb{B} \\ \textsf{stationary-distribution}\ N \Longleftrightarrow \\ \qquad N = \textsf{point}\ S\ \left(\lambda y.\ \displaystyle\int_x^+ \textsf{measure}\ (K\ x)\ \{y\}\ dN\right) \end{array}$$

It is often very easy to show that a Markov chain has a stationary distribution N: we only need to show that N is a solution to this equation system.

Example 30 (Stationary distributions of Fig. 1). For the essential class $\{C_1, C_2, C_3\}$ in Fig. 1 (a), we have the stationary distribution N with $\textsf{measure}\ N\ \{C_1\} = \textsf{measure}\ N\ \{C_2\} = 3/10$ and $\textsf{measure}\ N\ \{C_3\} = 4/10$. Fig. 1 (b) is an essential Markov chain its stationary distribution N is $\textsf{measure}\ N\ \{n\} = 1/2^{n+1}$.

Theorem 31 (Stationary distribution implies positive recurrence).

$$\begin{array}{l} \textsf{stationary-distribution}\ N \wedge \textsf{prob-space}\ N\ \wedge \\ \textsf{essential}\ C \wedge \textsf{measure}\ N\ (S - C) = 0 \wedge x \in C \implies \\ \textsf{pos-recurrent}\ x \quad \wedge \quad M\ x = 1/\textsf{measure}\ N\ \{x\} \end{array}$$

This lemma answers **Q3**, with the stationary distribution we compute $M\ x$.

Another property of the stationary distribution is, that it is the limit of the marginal distribution $p\ x\ y\ t$ when t goes to infinity. This limit does not always exist: a Markov chain where the graph is a cycle, and the transition probabilities are all 1 has no such limit. We introduce aperiodic classes to avoid these cases.

Definition 32 (Aperiodic classes).

$$\begin{array}{l} \textsf{aperiodic} \quad :: \quad \alpha\ \textsf{set} \Rightarrow \mathbb{B} \\ \textsf{aperiodic}\ C \Longleftrightarrow C \in S \big/ \textsf{communicating}\ \wedge \\ \qquad (\forall x \in C.\ \textsf{Gcd}\ \{i \mid 0 < i \wedge 0 < p\ x\ x\ i\} = 1) \end{array}$$

Lemma 33 (Aperiodic class implies non-zero p).

$$\textsf{aperiodic}\ C \implies \forall x \in C.\ \exists t.\ \forall t' \geq t.\ 0 < p\ x\ x\ t'$$

The stationary distribution is the limit of the marginal distributions.

Theorem 34 (Stationary distribution is asymptotic distribution).

$$\begin{array}{l} \textsf{stationary-distribution}\ N \wedge \textsf{prob-space}\ N \wedge \textsf{measure}\ N\ (S - C) = 0\ \wedge \\ \textsf{aperiodic}\ C \wedge \textsf{essential}\ C \wedge y \in C \implies \\ \displaystyle\int_x \Big| p\ x\ y\ t - \textsf{measure}\ N\ \{y\} \Big|\ d\textsf{count}\ C \xrightarrow{t \to \infty} 0, \\ x \in C \implies p\ x\ y\ t \xrightarrow{t \to \infty} \textsf{measure}\ N\ \{y\} \end{array}$$

[1] $N = \textsf{point}\ S\ f \Longleftrightarrow (\forall x \in S.\ \textsf{measure}\ N\ \{x\} = f\ x) \wedge \textsf{sets}\ N = \mathcal{P}(S)$

With this theorem we answer our final question **Q4**. The Fig. 1 (b) is aperiodic and has a stationary probability distribution, hence we only need to compare the values of the stationary distribution.

7 Discussion

Using measure theory to formalize Markov chains is not strictly necessary. Woess mentions in the introduction of [15] that it is also possible to consider the Markov chain on a time interval up to time t. Then the set of traces is countable and quantities can be expressed as (countable) sums. When an infinite time is involved one could take the limit over these quantities when the time t goes to infinity. However, applying measure theory provides a more generic framework: we can handle all cases as the involved functions and predicates are measurable. To handle measurability we even implemented a simple measurability prover.

In three other aspects measure theory was very helpful for us:

- In comparison to earlier formalizations of Markov chains we are not restricted to a finite state space. However, countable state spaces are more complicated when we sum over all states. Here we found it helpful to use the counting measure *count* S defined in Section 4.1, which allows us to use the Lebesgue integral as sum operator over an infinite set.
- The Lebesgue integral was also very helpful for the splitting rules (Lemma 2). As common way to reduce probabilities on Markov chains is the Chapman-Kolmogorov theorem: $p\ x\ y\ (n+m) = \sum_{z \in S} p\ x\ z\ n \cdot p\ z\ y\ m$. The splitting rules and our generic Chapman-Kolmogorov are much more flexible. Of course it is necessary to show measurability of f or P, but for this we use our simple measurability prover.
- It was important to us to show the equivalence between the stochastic-process and the transition-matrix interpretation of Markov chains. This was made possible with the measure space constructions in Section 4.1.

The last point strengthens that our formalization works on all Markov chains, and that we did not forget an assumption. For example, the formalization by Liu et al. [12] does not fulfill this guarantee. In their formalization of the time-homogeneous property they miss the assumption that the states are reachable at all. Hence the Markov chains they handle are restricted to Markov chains where the probability to reach a state is *always* non-zero, i.e. $\Pr(X\ t = x) \neq 0$.

The recent developments in analysis for Isabelle/HOL [6] helped us with the limit approaching from left and with translating limits from $\mathbb{R}$ to $\overline{\mathbb{R}}$. This was important when working with generating functions.

One restriction of our formalization is that the $\mathcal{P}_x$ assumes a starting state. We showed that it is possible to use an initial distribution by embedding the state space into the option type. However, a more direct approach would simplify this and we plan to change the definition of the trace space from $\mathcal{P}_x$ to $\mathcal{P}_I$ where I is the initial distribution on S.

8 Summary and Future Work

In this paper we formalized Markov chain theory. Compared to earlier formalizations we provide the following contributions:

- The Markov chain can be either a stochastic process or a transition matrix.
- The state space is allowed to be countably infinite.
- By using generating functions, we formalized the relations between different quantities and (positive) recurrence.
- Our formalization allows to verify positive recurrent states and the limit of the marginal distribution by using the stationary distribution.

Based on the work presented in this paper we want to formalize Markov decision processes. Such a formalization involves Markov chains with countably infinite state spaces, as the state encodes the history for the involved nondeterministic decisions.

References

1. Audebaud, P., Paulin-Mohring, C.: Proofs of randomized algorithms in Coq. S. of Comp. Prog. 74(8), 568–589 (2009), issue on Math. of Prog. Constr. (MPC 2006)
2. Cock, D.: Verifying probabilistic correctness in Isabelle with pGCL. In: Systems Software Verification. EPTCS, vol. 102, pp. 167–178 (2012)
3. Haftmann, F., Wenzel, M.: Local theory specifications in Isabelle/Isar. In: Berardi, S., Damiani, F., de'Liguoro, U. (eds.) TYPES 2008, LNCS, vol. 5497, pp. 153–168
4. Hansson, H., Jonsson, B.: A logic for reasoning about time and reliability. Tech. Rep. SICS/R90013, Swedish Institute of Computer Science (Dec 1994)
5. Hölzl, J.: Construction and Stochastic Applications of Measure Spaces in Higher-Order Logic. Ph.D. thesis, TU München (October 2012)
6. Hölzl, J., Immler, F., Huffman, B.: Type classes and filters for mathematical analysis in Isabelle/HOL. In: Interactive Theorem Proving (ITP 2013) (2013)
7. Hurd, J.: Formal Verification of Probabilistic Algorithms. Ph.D. thesis, University of Cambridge (2002)
8. Hurd, J., McIver, A., Morgan, C.: Probabilistic guarded commands mechanized in HOL. Theoretical Computer Science 346(1), 96–112 (Nov 2005)
9. Hölzl, J.: Theory files, `http://home.in.tum.de/hoelzl/cpp2013/`
10. Katoen, J.P., Zapreev, I.S., Hahn, E.M., Hermanns, H., Jansen, D.N.: The ins and outs of the probabilistic model checker MRMC. Perf. Eval. 68, 90–104 (2011)
11. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: CAV 2011. LNCS, vol. 6806, pp. 585–591 (2011)
12. Liu, L., Hasan, O., Aravantinos, V., Tahar, S.: Formal reasoning about classified markov chains in HOL. In: Interactive Theorem Proving (ITP 2013). LNCS
13. Popescu, A., Hölzl, J., Nipkow, T.: Formalizing probabilistic noninterference. Submitted to CPP 2013
14. Trivedi, K.S.: Probability & Statistics with Reliability, Queuing, and Computer Science Applications. Prentice-Hall (1982)
15. Woess, W.: Denumerable Markov Chains. European Mathematical Society (2009)