Formal Reasoning about Classified Markov Chains in HOL

Liya Liu, Osman Hasan, Vincent Aravantinos, and Sofiène Tahar

Dept. of Electrical & Computer Engineering, Concordia University
1455 de Maisonneuve W., Montreal, Quebec, H3G 1M8, Canada
{liy_liu,o_hasan,vincent,tahar}@ece.concordia.ca

Abstract. Classified Markov chains have been extensively applied to model and analyze various stochastic systems in many engineering and scientific domains. Traditionally, the analysis of these systems has been conducted using computer simulations and, more recently, also probabilistic model-checking. However, these methods either cannot guarantee accurate analysis or are not scalable due to the unacceptable computation times. As an alternative approach, this paper proposes to reason about classified Markov chains using HOL theorem proving. We provide a formalization of classified discrete-time Markov chains with finite state space in higher-order logic and the formal verification of some of their widely used properties. To illustrate the usefulness of the proposed approach, we present the formal analysis of a generic LRU (least recently used) stack model.

1 Introduction

In analyzing the stationary behaviors of Markovian models, it is quite common to categorize Markov chains into different classes depending on the properties exhibited by their states [3]. Some commonly used classes include *reducible*, *irreducible*, *periodic*, *aperiodic*, *regular* and *absorbing Markov chains*. Classified Markov chains are very useful for the dynamic analysis of systems as their properties allow us to judge long-run characteristics of Markovain systems, such as if a system will re-visit a particular state or to determine the time of the first visit to a state. Some of the widely used application areas of the classified Markov chains are reliability analysis, performance analysis and validation of models.

Traditionally, simulation is the most commonly applied computer-based analysis technique for Markovian systems. The main idea here is to utilize an equilibrium vector to approximate $vp_{ij}^{(n)}$, where v is any probability vector and $p_{ij}^{(n)}$ is the n-step transition probability. The main reason behind using the equilibrium vector is the high computational costs associated with $vp_{ij}^{(n)}$ for large values of n. Moreover, many rounding errors also creep into the analysis due to the involvement of computer arithmetic. Such approximations and inaccuracies pose a serious problem while analyzing highly sensitive and safety-critical applications.

Due to the extensive usage of Markov chains for safety-critical systems, probabilistic model checking has been recently proposed for analyzing Markovian

S. Blazy, C. Paulin-Mohring, and D. Pichardie (Eds.): ITP 2013, LNCS 7998, pp. 295–310, 2013.

systems. Probabilistic model checking tools, such as *PRISM* [15], *VESTA* [16] and *Ymer* [19], provide precise system analysis by modeling the stochastic behaviors, including its random components, using probabilistic state machines and exhaustively verifying their probabilistic properties. However, some algorithms implemented in these model checking tools are based on numerical methods. For example, the Power method [14], which is a well-known iterative method, is applied to compute the steady-state probabilities (or limiting probabilities) of Markov chains in PRISM. For this reason, most of the stationary properties analyzed in model checkers are time bounded. Moreover, probabilistic model checking often utilizes unverified algorithms and optimization techniques. Finally, model checking cannot be used to verify generic mathematical expressions for probabilistic analysis. In order to overcome these limitations, we proposed to use higher-order-logic theorem proving for analyzing Discrete Time Markov Chains (DTMCs) [10], where we presented a formal definition of DTMC and verified some of its properties using the HOL theorem prover. This formalization enabled us to formally analyze some simple Markovian models in HOL. In order to extend the capabilities of higher-order-logic theorem proving based analysis of Markovian models and thus be able to analyze a wider range of real-world systems, we present in the current paper the formalization of classified DTMCs.

In [8], the authors formally defined a time-homogeneous Markov chain based on the state space and the transition matrix in Isabelle/HOL, and they assumed no initial distribution or start state. Compared to their definition and the formalization presented in [10], which was based on the probability theory developed in [6], this paper describes a more generic higher-order-logic formalization of finite-state DTMC. Our work is based on a more general formalization of probability theory [11], which provides us with the flexibility to model inhomogeneous DTMCs or several random processes (involving DTMCs) containing distinct types of state spaces. We then build upon the formal DTMCs to formalize classified DTMCs and to formally verify the properties of aperiodic and irreducible DTMCs. For illustration purposes, we formally validate a least recently used (LRU) stack model using our formalization.

2 Formalization of DTMCs

A *probability space* is a measure space $(\Omega, \Sigma, \mathcal{P}r)$ such that $\mathcal{P}r(\Omega) = 1$ [3]. Σ is a collection of subsets of Ω (these should satisfy some closure axioms that we do not specify here) which are called *measurable sets*. In [12], a higher-order logic probability theory is developed, where given a probability space `p`, the functions `space` and `subsets` return the corresponding Ω and Σ, respectively. Mathematically, a *random variable* is a measurable function between a probability space and a *measurable space*, which refers to a pair $(S, \mathcal{A})$, where S is a set and $\mathcal{A}$ is a σ-algebra, i.e., a collection of subsets of S satisfying some particular properties [3]. In HOL, we write `random_variable X p s` to state that a function `X` is a random variable on a probability space `p` and the measurable outcome space `s`. Building on these foundations, measure theoretic formalizations of probability, Lebesgue integral and information theories are presented in [12].

A *stochastic process* [3] is a function $X : T \to \Omega$ where $T = \mathbb{N}$ (*discrete-time process*) or $T = \mathbb{R}$ (*continuous-time process*) and Ω is a measurable set called the *state space* of X. A *(finite-state) DTMC* is a discrete-time stochastic process that has a finite Ω and satisfies the *Markov property* [4]: for $0 \leq t_0 \leq \cdots \leq t_n$ and $f_0, \cdots, f_{n+1}$ in the state space, then: $\mathcal{P}r\{X_{t_{n+1}} = f_{n+1} | X_{t_n} = f_n, \ldots, X_{t_0} = f_0\} = \mathcal{P}r\{X_{t_{n+1}} = f_{n+1} | X_{t_n} = f_n\}$.

This allows to formalize the Markov property as follows:

Definition 1 (Markov Property).

```
⊢ ∀ X p s. mc_property X p s =
   (∀ t. random_variable (X t) p s) ∧
   ∀ f t n.
    increasing_seq t ∧ ℙ(⋂ₖ∈ [0,n−1]{x | X tₖ x = f k}) ≠ 0 ⇒
    (ℙ({x | X tₙ₊₁ x = f (n + 1)}|{x | X tₙ x = f n} ∩
                                   ⋂ₖ∈ [0,n−1]{x | X tₖ x = f k}) =
     ℙ({x | X tₙ₊₁ x = f (n + 1)}|{x | X tₙ x = f n}))
```

where `increasing_seq t` is defined as `∀ i j. i < j ⇒ t i < t j`, thus formalizing the notion of increasing sequence. The first conjunct indicates that the Markov property is based on a random process $\{X_t : \Omega \to S\}$. The quantified variable `X` represents a function of the random variables associated with time `t` which has the type `num`. This ensures the process is a *discrete time* random process. The random variables in this process are the functions built on the probability space `p` and a measurable space `s`. The conjunct $\mathbb{P}(\bigcap_{\mathtt{k}\in[0,\mathtt{n}-1]}\{\mathtt{x} \mid \mathtt{X}\ \mathtt{t_k}\ \mathtt{x} = \mathtt{f}\ \mathtt{k}\}) \neq 0$ ensures that the corresponding conditional probabilities are well-defined, where `f k` returns the $\mathtt{k}^{th}$ element of the state sequence.

A DTMC is usually expressed by specifying: an initial distribution p_0 which gives the probability of initial occurrence $\mathcal{P}r(X_0 = s) = p_0(s)$ for every state; and transition probabilities $p_{ij}(t)$ which give the probability of going from i to j for every pair of states i, j in the state space [13]. For states i, j and a time t, the *transition probability* $p_{ij}(t)$ is defined as $\mathcal{P}r\{X_{t+1} = j | X_t = i\}$, which can be easily generalized to *n-step transition probability*.

$$p_{ij}^{(n)} = \begin{cases} \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} & n = 0 \\ \mathcal{P}r\{X_{t+n} = j | X_t = i\} & n > 0 \end{cases}$$

This is formalized in HOL as follows Definition 2 below so that the discrete-time Markov chain (DTMC) can be formalized as Definition 3:

Definition 2 (Transition Probability).

```
⊢ ∀ X p s t n i j.
  Trans X p s t n i j =
  if n = 0 then
    if i ∈ space s ∧ j ∈ space s then
      if (i = j) then 1 else 0
```

```
    else 0
  else ℙ({x | X (t + n) x = j}|{x | X t x = i})
```

Definition 3 (DTMC).

```
⊢ ∀ X p s Linit Ltrans. dtmc X p s Linit Ltrans =
  mc_property X p s ∧ (∀ i. i ∈ space s ⇒ {i} ∈ subsets s) ∧
  ∀ i. i ∈ space s ⇒ (Linit i = ℙ{x | X t x = i}) ∧
  ∀ t i j. (ℙ{x | X t x = i} ≠ 0) ⇒
           (Ltrans t i j = Trans X p s t 1 i j)
```

It is important to note that `X` is polymorphic, i.e., it is not constrained to a particular type, which is a very useful feature of our definition.

In practice, many applications actually make use of *time-homogenous DTMCs*, i.e., DTMCs with finite state-space and time-independent transition probabilities [2]. They can be formalized as follows:

Definition 4 (Time homogeneous DTMC).

```
⊢ ∀ X p s p₀ pᵢⱼ. th_dtmc X p s p₀ pᵢⱼ =
   dtmc X p s p₀ pᵢⱼ ∧ FINITE (space s) ∧
   ∀ t i j. Trans X p s (t + 1) 1 i j = Trans X p s t 1 i j
```

For time-homogenous DTMCs, $\forall\ t\ t'.\ p_{ij}(t) = p_{ij}(t')$ and thus $p_{ij}(t)$ is simply written as p_{ij}.

It is often the case that we are interested in the probability of some specific states as time tends to infinity under certain conditions. This is the main reason why stationary behaviors of stochastic processes are frequently analyzed in engineering and scientific domains. There is no exception for Markovian systems.

Let $\{X_t\}_{t\geq 0}$ be a Markov chain having state space Ω and transition probability p_{ij} for going from a state with value i to a state with value j. If $\pi(i)$, $i \in \Omega$, are nonnegative numbers summing to one, and if $j \in \Omega$, then $\pi(j) = \sum_{i\in\Omega} \pi(i)p_{ij}$ is called a *stationary distribution*. The corresponding HOL definition is as follows.

Definition 5 (Stationary Distribution).

```
⊢ ∀ p X f s. stationary_dist p X f s =
  ∑ (f k) = 1 ∧
  ∀ i. i ∈ space s ⇒
       0 < f i ∧ ∀ t. f i = ∑ f k * Trans X p s t 1 k i
```

Using these fundamental definitions, we formally verified most of the classical properties of DTMCs with finite state-space in HOL. Some of the relevant ones to the context of this paper are presented later.

3 Formalization of Classified DTMCs

In this section, we first formalize some foundational notions of classified Markov chains. Then, we use these results along with our formal definition of DTMC to formalize classified Markov chains. The foremost concept of states classification is the *first passage time* τ_j, or the *first hitting time*, which is defined as the

minimum time required to reach a state j from the initial state i:

$$\tau_j = min\{t > 0 : X_t = j\}.$$

The first passage time can be defined in HOL as:

Definition 6 (First Passage Time).

```
⊢ FPT X x j = MINSET {t | 0 < t ∧ (X t x = j)}
```

where `X` is a random process and `x` is a sample in the probability space associated with the random variable X_t. Note that the first passage time is also a random variable.

The conditional distribution of τ_j defined as the probability of the events starting from state i and visiting state j at time n is expressed as $f_{ij}^{(n)} = \mathcal{P}r\{\tau_j = n|X_0 = i\}$. It can be formalized in HOL as follows:

Definition 7 (Probability of First Passage Events).

```
⊢ f X p i j n = ℙ({x | FPT X x j = n}|{x | X 0 x = i})
```

Another important notion, denoted as f_{ij}, is the probability of the events starting from state i and visiting state j at all times n, is expressed as $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$. It can be expressed in HOL as `(λ n. f X p i j n) sums` f_{ij}. Another interesting concept is f_{jj}, which provides the probability of events starting from state j and eventually returning back to j. If $f_{jj} = 1$, then the *mean return time* of state j is defined as $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$. The existence of this infinite summation can be specified as `summmable (λ n. n * f X p j j n)` in HOL.

A state j in a DTMC $\{X_t\}_{t\geq 0}$ is called *transient* if $f_{jj} < 1$, and *persistent* if $f_{jj} = 1$. If the mean return time μ_j of a persistent state j is finite, then j is said to be *persistent nonnull state* (or *positive persistent state*). Similarly, if μ_j is infinite, then j is termed as *persistent null state*.

The greatest common divisor (*gcd*) of a set is a frequently used mathematical concept in defining classified states. We formalize the gcd of a set as follows:

Definition 8 (gcd of a Set).

```
⊢ GCD_SET A = MAXSET {r | ∀ x. x ∈ A ⇒ divides r x}
```

For a state j, a *period* of j is any n such that $p_{jj}^{(n)}$ is greater than 0. We write $d_j = gcd\ \{n : p_{jj}^{(n)} > 0\}$ as the gcd of the set of all periods.

A state i is said to be *accessible* from a state j (written $j \rightarrow i$), if the n-step transition probability of the events from state i to j is nonzero. Two states i, j are called *communicating states* (written $i \leftrightarrow j$) if they are mutually accessible. A state j is an *absorbing state* if $p_{jj} = 1$. The formalization of some other foundational notions of classified states is given in Table 1. Now, we build upon the above mentioned definitions to formalize classified DTMCs. Usually, a DTMC is said to be *irreducible* if every state in its state space can be reached from any other state including itself in finite steps.

Table 1. Formalization of Classified States

Definition	Condition	HOL Formalization
Transient State	$f_{jj} < 1$	⊢ ∀ X p j. Transient_state X p j = ∀ x. {t | 0 < t ∧ (X t x = j)} ≠ ∅ ∧ (∃ s. s < 1 ∧ (λ n. f X p j j n) sums s)
Persistent State	$f_{jj} = 1$	⊢ ∀ X p j. Persistent_state X p j = ∀ x. {t | 0 < t ∧ (X t x = j)} ≠ ∅ ∧ (λ n. f X p j j n) sums 1
Persistent Nonnull State	$f_{jj} = 1$ $\mu_j < \infty$	⊢ ∀ X p j. Nonnull_state X p j = Persistent_state X p j ∧ summable (λ n. n * f X p j j n)
Persistent Null State	$f_{jj} = 1$ $\mu_j = \infty$	⊢ ∀ X p j. Null_state X p j = Persistent_state X p j ∧ ∼ summable (λ n. n * f X p j j n)
Periods of a State	$0 < \mathrm{n}$ $0 < p^n_{jj}$	⊢ ∀ X p s j. Period_set X p s j = {n | 0 < n ∧ ∀ t. 0 < Trans X p s t n j j}
GCD of a Period Set	d_j	⊢ ∀ X p s j. Period X p s j = GCD_SET (Period_set X p s j)
Periodic State	$d_j > 1$	⊢ ∀ X p s j. Periodic_state X p s j = (1 < Period X p s j) ∧ (Period_set X p s j ≠ ∅)
Aperiodic State	$d_j = 1$	⊢ ∀ X p s j. Aperiodic_state X p s j = (Period X p s j = 1) ∧ (Period_set X p s j ≠ ∅)
Accessibility	$\mathrm{i} \to \mathrm{j}$	⊢ ∀ X p s i j.Accessibility X p s i j = ∀ t. ∃ n. 0 < n ∧ 0 < Trans X p s t n i j
Communicating State	$\mathrm{i} \leftrightarrow \mathrm{j}$	⊢ ∀ X p s i. Communicating_states X p s i j = (Accessibility X p s i j) ∧ (Accessibility X p s j i)
Absorbing State	$p_{jj} = 1$	⊢ ∀ X p s j. Absorbing_states X p s j = (Trans X p s t 1 j j = 1)

Definition 9 (Irreducible DTMC).

```
⊢ Irreducible_mc X p s p0 pij =
   th_dtmc X p s p0 pij ∧
  (∀ i j. i ∈ space s ∧ j ∈ space s ⇒
          Communicating_states X p s i j)
```

whereas if there exists a state in the state space of a DTMC, which cannot reach some other states, then this DTMC is called *reducible*.

Definition 10 (Reducible DTMC).

```
⊢ Reducible_mc X p s p0 pij =
  th_dtmc X p s p0 pij ∧
  (∃ i j. i ∈ space s ∧ j ∈ space s ∧
         ∼ Communicating_states X p s i j)
```

A DTMC is considered as *aperiodic* if every state in its state space is an aperiodic state; otherwise it is a *periodic DTMC*.

Definition 11 (Aperiodic DTMC).

```
⊢ Aperiodic_mc X p s p0 pij =
  th_dtmc X p s p0 pij ∧
  ∀ i. i ∈ space s ⇒ Aperiodic_state X p s i
```

Definition 12 (Periodic DTMC).

```
⊢ Periodic_mc X p s p0 pij =
  th_dtmc X p s p0 pij ∧
  ∃ i. i ∈ space s ∧ Periodic_state X p s i
```

If at least one absorbing state exists in a DTMC and it is possible to go to the absorbing state from every non-absorbing state, then such a DTMC is named as *absorbing DTMC*.

Definition 13 (Absorbing DTMC).

```
⊢ Absorbing_mc X p s p0 pij =
  th_dtmc X p s p0 pij ∧
  ∃ i. i ∈ space s ∧ Absorbing_state X p s i ∧
       ∀ j. j ∈ space s ⇒ Communicating_state X p s i j
```

Finally, if there exists some n such that $p_{ij}^{(n)} > 0$ for all states i and j in a DTMC, then this DTMC is defined as a *regular DTMC*.

Definition 14 (Regular DTMC).

```
⊢ Regular_mc X p s p0 pij =
  th_dtmc X p s p0 pij ∧
  ∃ n. ∀ i j. i ∈ space s ∧ j ∈ space s ⇒
               Trans X p s t n i j > 0
```

To the best of our knowledge, the above mentioned definitions constitute the first formalization of classified DTMCs in higher-order logic. Their main utility is to formally specify and analyze the dynamic features of Markovian systems within a sound theorem prover as will be demonstrated in Section 5.

4 Verification of DTMC Properties

In this section, we utilize the definitions given above, to verify some of the most frequently used properties of DTMCs and classified DTMCs. The formal verification of these properties not only ensure the correctness of our definitions but also plays a vital role in formal reasoning about DTMCs and classified DTMCs in a theorem prover.

4.1 DTMC Properties

The *joint probability distribution* of a DTMC is the probability of a chain of states to occur. It is very useful in analyzing multi-stage experiments. In addition, this concept is the basis for the frequently used joint probability generating functions.

Theorem 1 (Joint Probability Distribution).

A joint probability distribution of n discrete random variables $X_0, \ldots X_n$ *in a finite DTMC* $\{X_t\}_{t\geq 0}$ *satisfies:*

$$\mathcal{P}r(X_t = L_0, \cdots, X_{t+n} = L_n) = \prod_{k=0}^{n-1} \mathcal{P}r(X_{t+k+1} = L_{k+1} | X_{t+k} = L_k) \mathcal{P}r(X_t = L_0)$$

```
⊢ ∀ X p s p L p₀ pᵢⱼ n.
   dtmc X p s p₀ pᵢⱼ ⇒
   ℙ(⋂_{k=0}^{n} {x | X (t + k) x = EL k L}) =
   ∏_{k=0}^{n-1} ℙ({x | X (t + k + 1) x = EL (k + 1) L} |
          {x | X (t + k) x = EL k L}) ℙ{x | X t x = EL 0 L}
```

The proof of Theorem 1 is based on induction on the variable n, Definition 3 and some arithmetic reasoning.

The Chapman-Kolmogorov equation [3] is a widely used property of time homogeneous DTMCs. It basically gives the probability of going from state i to j in $m + n$ steps. Assuming the first m steps take the system from state i to some intermediate state k and the remaining n steps then take the system from state k to j, we can obtain the desired probability by adding the probabilities associated with all the intermediate steps.

Theorem 2 (Chapman-Kolmogorov Equation).

For a finite time homogeneous DTMC $\{X_t\}_{t\geq 0}$*, its transition probabilities satisfy the Chapman-Kolmogorov Equation*

$$p_{ij}^{(m+n)} = \sum_{k\in\Omega} p_{ik}^{(m)} p_{kj}^{(n)}$$

```
⊢ ∀ X p s i j t m n p₀ pᵢⱼ.
   th_dtmc X p s p₀ pᵢⱼ ⇒
   Trans X p s t (m + n) i j =
   ∑_{k∈space s} (Trans X p s t m i k * Trans X p s t n k j)
```

The proof of Theorem 2 again involves induction on the variables m and n and both of the base and step cases are discharged using the following lemma:

Lemma 1 (Multistep Transition Probability).

```
⊢ ∀ X p s i j t m p₀ pᵢⱼ.
   th_dtmc X p s p₀ pᵢⱼ ⇒
   Trans X p s t (m + 1) i j =
   ∑_{k∈space s} (Trans X p s t 1 k j * Trans X p s t m i k)
```

The proof of Lemma 1 is primarily based on Definitions 3 and 4 and the additivity property of probabilities.

The unconditional probabilities associated with a Markov chain are called *absolute probabilities*, which can be computed by applying the initial distributions and n-step transition probabilities. From now, let us write $p_i^{(n)}$ for the probability $\mathcal{P}r(X_n = j)$. We then have the following result:

Theorem 3 (Absolute Probability).

In a finite time homogeneous DTMC, the absolute probabilities $p_j^{(n)}$ *satisfy*

$$p_j^{(n)} = \mathcal{P}r(X_n = j) = \sum_{k \in \Omega} \mathcal{P}r(X_0 = k)\mathcal{P}r(X_n = j | X_0 = k)$$

```
⊢ ∀ X p s j n p₀ pᵢⱼ.
   th_dtmc X p s p₀ pᵢⱼ ⇒
   ℙ{x | X n x = j} =
   ∑_{k∈space s} ℙ{x | X 0 x = k}ℙ({x | X n x = j}|{x | X 0 s = k})
```

The proof of Theorem 3 is based on the Total Probability theorem along with some basic arithmetic and probability theoretic reasoning.

The formal proof script for the above mentioned properties and many other useful properties is available at [9].

4.2 Classified DTMC Properties

Among the classified DTMCs formalized in the previous section, `aperiodic and irreducible DTMCs` are considered to be the most widely used ones in analyzing Markovian systems because of their attractive stationary properties, i.e., their limit probability distributions are independent of the initial distributions. For this reason, we now focus on the verification of some key properties of aperiodic and irreducible DTMCs [5].

Theorem 4 (Closed Period Set).

In an aperiodic DTMC, the set of the times when state i has a non-null probability of being visited is closed under addition.

```
⊢ ∀ X p s p₀ pᵢⱼ i.
   Aperiodic_DTMC X p s p₀ pᵢⱼ ∧ i ∈ space s ⇒
   ∀ a b. a ∈ Period_set X p s i ∧ b ∈ Period_set X p s i ⇒
          (a + b) ∈ Period_set X p s i
```

We verified the above theorem by using Theorem 2 and arithmetic and set theoretic reasoning.

Another key property of an aperiodic DTMC states that the transition probability $p_{ij}^{(n)}$ is greater than zero, for all states i and j in its state space, after n steps. It is very useful in analyzing the stability or reliability of real-world systems.

Theorem 5 (Positive Return Probability).

For any state i in the finite state space S of an aperiodic DTMC, there exists an $N < \infty$ *such that* $0 < p_{ii}^{(n)}$, *for all* $n \geq N$.

```
⊢ ∀ X p s p₀ pᵢᵢ i t.
   Aperiodic_DTMC X p s p₀ pᵢᵢ ∧ i ∈ space s ⇒
   ∃ N. ∀ n. N ≤ n ⇒ 0 < Trans X p s t n i i
```

The formal reasoning about the correctness of the above theorems involves Theorems 2 and 4 and the following lemmas, along with some arithmetic reasoning and set theoretic reasoning.

Lemma 2 (Positive Element in a Closed Set).

If an integer set S contains at least one nonzero element and S is closed under addition and subtraction, then $S = \{kc; k \in \mathbb{Z}\}$, where c is the least positive element of S.

```
⊢ ∀ s:int → bool. s ≠ ∅ ∧
   (∀ a b. a ∈ s ∧ b ∈ s ⇒ (a + b) ∈ s ∧ (a - b) ∈ s) ⇒
   0 < MINSET {r | 0 < r ∧ r ∈ s} ∧
   (s = {r | ?k. r = k * MINSET {r | 0 < r ∧ r ∈ s}})
```

Lemma 3 (Linearity of Two Integer Sequences).

For a positive integer sequence $a_1, a_2, \cdots, a_k$, there exists an integer sequence $n_1, n_2, \cdots, n_k$, such that $d = \sum_{i=1}^{k} n_i a_i$, where d is the greatest common divisor of sequence $a_1, a_2, \cdots, a_k$.

```
⊢ ∀ a k. 0 < k ∧ (∀ i. i ≤ k ⇒ 0 < a i) ⇒
   (∃ n. GCD_SET {a i | i ∈ [0, k]} = ∑ᵢ₌₀ᵏ n i * a i)
```

Lemma 4 (Least Number).

If a set of positive integers A is nonlattice, i.e., its gcd is 1, and closed under addition, then there exists an integer $N < \infty$ such that $n \in A$ for all $N \leq n$.

```
⊢ ∀ (A:int → bool) a.
   (A = {a i | 0 < a i ∧ i ∈ UNIV(:num)}) ∧ (GCD_SET A = 1) ∧
   (∀a b. a ∈ A ∧ b ∈ A ⇒ (a + b) ∈ s) ⇒ (∃N. {n | N ≤ n} ⊂ A)
```

The proofs of Lemmas 2, 3 and 4 are based upon various summation properties of integer sets. These properties are not available in the HOL libraries and thus had to be verified as part of our development.

Theorem 6 (Existence of Positive Transition Probabilities).

For any aperiodic and irreducible DTMC with finite state space S, there exists an N, for all $n \geq N$, such that the n-step transition probability $p_{ij}^{(n)}$ is non-zero, for all states i and $j \in S$.

```
⊢ ∀ X p s p₀ pᵢⱼ i j t.
   Aperiodic_DTMC X p s p₀ pᵢⱼ ∧ Irreducible_DTMC X p s p₀ pᵢⱼ ∧
   i ∈ space s ∧ j ∈ space s ⇒
   ∃ N. ∀ n. N ≤ n ⇒ 0 < Trans X p s t n i j
```

We proceed with the proof of Theorem 6 by performing case analysis on the equality of i and j. The rest of the proof is primarily based on Theorems 2 and 5, Definition 1 and Lemmas 3 and 4.

Theorem 7 (Existence of Long-run Transition Probabilities).

For any aperiodic and irreducible DTMC with finite state space S and transition probabilities p_{ij}, there exists $\lim_{n\to\infty} p_{ij}^{(n)}$, for all states i and $j \in S$.

```
⊢ ∀ X p s p₀ pᵢⱼ i j t.
   Aperiodic_DTMC X p s p₀ pᵢⱼ ∧ Irreducible_DTMC X p s p₀ pᵢⱼ ⇒
   ∃ u. (λ n. Trans X p s t n i j → u)
```

We firstly prove the monotonic properties of M_j^n and m_j^n, which are the maximum and minimum values of the set $\{n \leq 1: p_{ij}^{(n)}>0\}$, respectively. Then, the proof is completed by verifying the convergence of the sequence $(M_j^n$ - $m_j^n)$ for all n by applying Theorem 2 and some properties of real sequences. It is important to note that we do not need to use the assumption $j \in$ *space* s here, like all the other theorems, as $\forall\ n\ j.\ j \notin$ *space* $s \Rightarrow (p_j^{(n)} = 0)$, which in turn implies $\lim_{n\to\infty} p_j^{(n)} = 0$ and $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. The long-run probability distributions are often considered in the convergence analysis of random variables in stochastic systems. It is not very easy to verify that the limit probability distribution of a certain state exists in a generic non-trivial DTMC, because the computations required in such an analysis are often tremendous. However, in the aperiodic and irreducible DTMCs, we can prove that all states possess limiting probability distribution, by the following two theorems.

Theorem 8 (Existence of Long-run Probability Distributions).

For any aperiodic and irreducible DTMC with finite state space S, there exists $\lim_{n\to\infty} p_i^{(n)}$, for any state $i \in S$.

```
⊢ ∀ X p s p₀ pᵢⱼ i.
   Aperiodic_DTMC X p s p₀ pᵢⱼ ∧ Irreducible_DTMC X p s p₀ pᵢⱼ ⇒
   ∃ u. (λ n. ℙ{x | X n x = i} → u)
```

We used Theorems 3 and 7, along with some properties of the limit of a sequence, to prove this theorem in HOL.

Theorem 9 (Existence of Steady State Probability).

For every state i in an aperiodic and irreducible DTMC, $\lim_{n\to\infty} p_i^{(n)}$ is a stationary distribution.

```
⊢ ∀ X p s p₀ pᵢⱼ.
   Aperiodic_DTMC X p s p₀ pᵢⱼ ∧ Irreducible_DTMC X p s p₀ pᵢⱼ ⇒
    (stationary_dist p X (λ i. lim ℙ{x | X n x = i}) s)
                              n→∞
```

The proof of Theorem 9 involves rewriting with Definition 5 and then splitting it into the following three subgoals:

- $0 \leq \lim_{n\to\infty} p_j^{(n)}$
- $\sum_{i\in\Omega} \lim_{n\to\infty} p_i^{(n)} = 1$
- $\lim_{n\to\infty} p_j^{(n)} = \sum_{i\in\Omega} \lim_{n\to\infty} p_i^{(n)} p_{ij}$

Utilizing the probability bounds theorem, we can prove the first subgoal. The proof of the second subgoal is primarily based on the additivity property of conditional probability [7]. Then the last subgoal can be proved by applying the linearity of limit of a sequence and the linearity of real summation.

All theorems presented in this section would facilitate the formal reasoning about the system properties that can be modeled using classified Markov chains. For illustration purposes, we present the formal analysis of a LRU stack model in the next section.

5 Formal Validation of LRU Stack Model

With the rapid development of computer technology, cache memory management becomes indispensable in computer architectures. The memory reference behaviors of various programs is one of the main deciding factors in designing efficient virtual memory operating systems. The Least Recently Used (LRU) stack model describes a behavior of reference strings where the probability of referencing a given page i at time t depends on the pages referenced in the closest past. In [1], the authors assumed the distance string for referencing a page as a sequence of independent identically distributed (IID) random variables in their LRU stack model, which was described as an aperiodic and irreducible DTMC [18]. However, in [17], the authors argued that the model constructed in [1] was not able to correctly depict the behavior of the LRU algorithm in multiprogramming. We want to formally verify the results of the latter authors using our formalization of aperiodic and irreducible DTMC.

5.1 LRU Stack Model

In a Least Recently Used (LRU) stack model, as shown in Figure 1, a sequence of stacks $s_1 s_2 \cdots s_t \cdots$ are associated with a reference string $w = \mathrm{x}_1\mathrm{x}_2\cdots\mathrm{x}_{t+1}\cdots$. Any stack s_t is a n-tuple $(j_1, j_2, \cdots, j_n)$, where j_i refers to the ith most recently referenced page at time t [18]. Let D_t be the position of the page x_t in the stack s_{t-1}. Then the distance string is $\mathrm{D}_1\mathrm{D}_2\cdots\mathrm{D}_t\cdots$, which is associated with the referencing string. This distance string can be modeled as a sequence of independent and identically distributed (IID) random variables [1], which makes their probability mass function (PMF) as $\mathcal{P}r(D_t = i) = \mathrm{a}_i$, where $i = 1, 2, \ldots,$ n and refers to the position of the least recently used page in the stack at time t, and $\sum_{j=1}^{n} \mathrm{a}_j = 1$. This way the distribution function becomes $\mathcal{P}r(D_t \leq i) = \sum_{j=1}^{i} \mathrm{a}_j$. If a tagged page occupies the ith position in the stack at time t, which

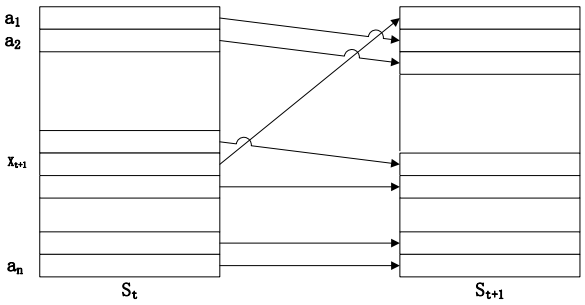

Fig. 1. LRU Stack updating Procedure [18]

is expressed as s_t in Figure 1, then the position of this page in the stack s_{t+1} depends on the next reference x_{t+1} and the position of this page in the stack s_t.

Based on the described updating procedure in the LRU stack, the evaluation of the page-fault rate of the LRU paging algorithm becomes quite simple. If the evaluated program has been allocated i page frames of main memory, then a page fault will occur at time t when $D_t > i$. Hence, the page fault probability is

$$\mathcal{F}(LRU) = \mathcal{P}r(D_t > i) = 1 - \sum_{j=1}^{i} a_j$$

The movement of the tagged page through the LRU state is then a random process $\{E_t\}_{t\geq 0}$. If the page occupies the ith position in stack s_t, then $E_t = i$, for all i, $1 \leq i \leq n$. Now, we have the following transition probabilities [18]:

$$p_{i1} = \mathcal{P}r(E_{t+1} = 1 | E_t = i) = \mathcal{P}r(D_{t+1} = i) = a_i, 1 \leq i \leq n$$

$$p_{ii} = \mathcal{P}r(E_{t+1} = i | E_t = i) = \mathcal{P}r(D_{t+1} < i) = \sum_{j=1}^{i-1} a_{j-1}, 2 \leq i \leq n$$

$$p_{i,i+1} = \mathcal{P}r(E_{t+1} = i+1 | E_t = i) = \mathcal{P}r(D_{t+1} > i) = 1 - \sum_{j=1}^{i} a_{j-1}, 1 \leq i \leq n-1$$

$$p_{i,j} = 0, otherwise.$$

The LRU stack is then described as an aperiodic and irreducible DTMC by assuming $a_i > 0$ for all $i \in [1, n]$ [18]. The state diagram of this aperiodic and irreducible DTMC is shown in Figure 2, where we can find that the transition probabilities can be expressed as the following higher-order logic function [18]:

Definition 15 (Transition Probability Matrix).

```
⊢ Lt a t i j =
  if (j = 1) then a i else
  if (j - i = 1) then 1 - ∑ i j=1 a j else
  if (j = i) then ∑ i−1 j=1 a j else 0
```

which can be used to formalize the LRU stack model as:

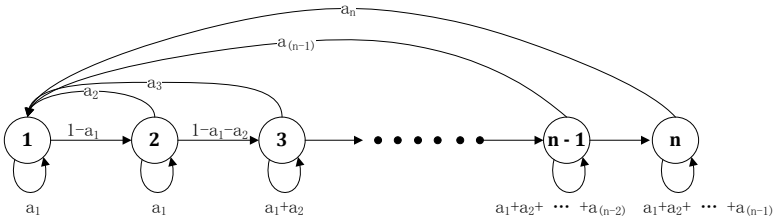

Fig. 2. The State Diagram for the LRU Stack Model

Definition 16 (LRU Model).

```
⊢ LRU_model X p a n p0.
  Aperiodic_DTMC X p ([1, n], POW ([1, n])) p0 (Lt a) ∧
  Irreducible_DTMC X p ([1, n], POW ([1, n])) p0 (Lt a) ∧
  1 <= n ∧ (∀ j. 0 < j ∧ j ≤ n ⇒ 0 < a j) ∧ (∑_{j=1}^{n} a j = 1)
```

where the state space is described as a pair `([1, n], POW ([1, n]))`, in which the first element contains all the states $\{1, 2, \cdots, n\}$ and the second one is the sigma algebra of the first element. The condition $(1 \leq n)$ is used to avoid the case when the length of the referencing string is zero. The other two conditions represent the specification of the model mentioned above.

5.2 Verification of the Property

Using the formal definition of this LRU stack model, we can now formally reason about its limiting distributions, which are mainly used to describe the stationary behaviors of this model.

Theorem 10 (Existence of LRU in the Limiting State Distribution).

In the LRU stack model, there exists $\lim_{t\to\infty} p_i^{(n)}$*, for every i ∈ [1, n].*

```
⊢ ∀ X p a n p0 i.
  LRU_model X p a n p0 ∧ i ∈ [1, n] ⇒
  ∃ u. (λ t. ℙ{x | X t x = i} → u)
```

We verified this property by directly applying Theorem 8 and the definition of limit of a real sequence.

Theorem 11 (LRU Stationary Limiting State Distribution).

In the LRU stack model, $\lim_{t\to\infty} p_i^{(n)} = \frac{1}{n}$*, for every i ∈ [1, n].*

```
⊢ ∀ X p a n p0 i.
  LRU_model X p a n p0 ∧ i ∈ [1,n] ⇒ lim_{t→∞} ℙ{x | X t x = i} = 1/n
```

The proof of this property is primarily based on Theorems 3 and 11 along with the following lemma:

Lemma 5 (Identity Limiting State Distribution).

$$\vdash \forall\ \texttt{X p a n p}_0\ \texttt{i j}.\ \texttt{LRU\_model X p a n p}_0 \wedge \texttt{i} \in \texttt{[1, n]} \wedge \texttt{j} \in \texttt{[1, n]} \Rightarrow \lim_{\texttt{t}\to\infty}\mathbb{P}\{\texttt{x | X t x = i}\} = \lim_{\texttt{t}\to\infty}\mathbb{P}\{\texttt{x | X t x = j}\}$$

The HOL proof of the above lemma is based on Theorem 3 along with some arithmetic reasoning.

Theorem 11 implies that $\lim_{\texttt{t}\to\infty} p_i^{(n)}$ (for any tag i) is independent of its initial distribution and the position of the tagged page has an equal probability to be in any stack position. This means that any page is equally likely to be referenced in the long run. As a result, it concludes that this LRU stack specification does not cover the case of nonuniform page referencing behavior of programs. Thus, we have been able to formally verify the numerical methods result presented in [17].

The HOL code developed for the formalization and verification of the classified DTMCs is totally around 8000 lines and the proof script for verifying Theorems 11 and 10 is about 300 lines long, which are available at [9]. The ability to formally verify theorems involving classified Markovian models and the short script clearly indicates the usefulness of the formalization, presented in the earlier section of the paper, as without them the reasoning could not have been done in such a straightforward manner.

6 Conclusion

This paper presents the formalization of classified DTMCs along with some important prerequisites related to the formalization of DTMCs with finite state-space in a higher-order logic theorem prover. Our results facilitate the formal analysis of classified DTMCs and provides the foundations for formalizing more advanced concepts of Markov chain theory, like hidden Markov chains, Markov decision process and other useful properties. Due to the inherent soundness of theorem proving, our work guarantees to provide accurate results, which is a very useful feature while analyzing stationary behaviors of a system associated with safety or mission-critical systems. In order to illustrate the usefulness of the proposed approach, we formally analyzed a LRU stack model using the definitions of aperiodic and irreducible DTMC and their formally verified properties. Our results exactly matched the conclusion and the corresponding experimental results in [17], which ascertains the precise nature of the proposed approach.

The presented work opens the door to a new and very promising research direction, i.e., integrating HOL theorem proving in the domain of analyzing classified DTMCs. We are currently working on extending the set of formally verified properties regarding DTMCs and extending our work to time-inhomogeneous discrete-time Markov chains and Markov Decision Process (MDP), which will enable us to formally analyze a wider range of systems. We also plan to build upon the formalization of continuous random variables and statistical properties to formalize Continuous-Time Markov Chains (CTMC) to be able to formally reason about statistical characteristics of a broader scope of Markovian models.

References

1. Avi-Itzhak, B., Heyman, D.P.: Approximate Queuing Models for Multiprogramming Computer Systems. Operations Research 21(6), 1212–1230 (1973)
2. Baier, C., Katoen, J.: Principles of Model Checking. MIT Press (2008)
3. Bhattacharya, R.N., Waymire, E.C.: Stochastic Processes with Applications. John Wiley & Sons (1990)
4. Chung, K.L.: Markov chains with stationary transition probabilities. Springer (1960)
5. Häggström, O.: Finite Markov Chains and Algorithmic Applications. Cambridge University Press (2002)
6. Hasan, O.: Formal Probabilistic Analysis using Theorem Proving. PhD Thesis, Concordia University, Montreal, QC, Canada (2008)
7. Hasan, O., Tahar, S.: Reasoning about Conditional Probabilities in a Higher-Order-Logic Theorem Prover. Journal of Applied Logic 9(1), 23–40 (2011)
8. Hölzl, J., Nipkow, T.: Interactive verification of markov chains: Two distributed protocol case studies. In: Quantities in Formal Methods, EPTCS, pp. 17–31 (2012)
9. Liu, L. (2013), `http://hvg.ece.concordia.ca/code/hol/cdtmc/`
10. Liu, L., Hasan, O., Tahar, S.: Formalization of Finite-State Discrete-Time Markov Chains in HOL. In: Bultan, T., Hsiung, P.-A. (eds.) ATVA 2011. LNCS, vol. 6996, pp. 90–104. Springer, Heidelberg (2011)
11. Mhamdi, T.: Information-Theoretic Analysis using Theorem Proving. PhD Thesis, Concordia University, Montreal, QC, Canada (2012)
12. Mhamdi, T., Hasan, O., Tahar, S.: On the Formalization of the Lebesgue Integration Theory in HOL. In: Kaufmann, M., Paulson, L.C. (eds.) ITP 2010. LNCS, vol. 6172, pp. 387–402. Springer, Heidelberg (2010)
13. Norris, J.R.: Markov Chains. Cambridge University Press (1999)
14. Parker, D.A.: Implementation of Symbolic Model Checking for Probabilitics Systems. PhD Thesis, University of Birmingham, Birmingham, UK (2002)
15. PRISM (2013), `http://www.prismmodelchecker.org`
16. Sen, K., Viswanathan, M., Agha, G.: VESTA: A Statistical Model-Checker and Analyzer for Probabilistic Systems. In: IEEE International Conference on the Quantitative Evaluation of Systems, pp. 251–252 (2005)
17. Shedler, G., Tung, C.: Locality in page reference strings. SIAM Journal on Computing 1(3), 218–241 (1972)
18. Trivedi, K.S.: Probability and Statistics with Reliability, Queuing, and Computer Science Applications. John Wiley & Sons (2002)
19. YMER (2013), `http://www.tempastic.org/ymer/`