Approximate Verification of the Symbolic Dynamics of Markov Chains

Manindra Agrawal
School of Electrical and Computer Engineering
Indian Institute of Technology
Kanpur, India 208016.
Email: manindra@iitk.ac.in

S. Akshay
IRISA, ENS Cachan Bretagne
Campus de Beaulieu
Rennes, France 35042.
Email: akshay@irisa.fr

Blaise Genest
CNRS, UMI IPAL
joint with NUS
and A*STAR/I2R
Singapore 138632.
Email: bgenest@irisa.fr

P. S. Thiagarajan
School of Computing
National University of Singapore
Singapore 117417.
Email: thiagu@comp.nus.edu.sg

***Abstract*—A finite state Markov chain M is often viewed as a probabilistic transition system. An alternative view - which we follow here - is to regard M as a linear transform operating on the space of probability distributions over its set of nodes. The novel idea here is to discretize the probability value space [0,1] into a finite set of intervals. A concrete probability distribution over the nodes is then symbolically represented as a tuple D of such intervals. The i-th component of the discretized distribution D will be the interval in which the probability of node i falls.**

The set of discretized distributions is a finite set and each trajectory, generated by repeated applications of M to an initial distribution, will induce a unique infinite string over this finite set of letters. Hence, given a set of initial distributions, the symbolic dynamics of M will consist of an infinite language L over the finite alphabet of discretized distributions. We investigate whether L meets a specification given as a linear time temporal logic formula whose atomic propositions will assert that the current probability of a node falls in an interval.

Unfortunately, even for restricted Markov chains (for instance, irreducible and aperiodic chains), we do not know at present if and when L is an (omega)-regular language. To get around this we develop the notion of an epsilon-approximation, based on the transient and long term behaviors of M. Our main results are that, one can effectively check whether (i) for each infinite word in L, at least one of its epsilon-approximations satisfies the specification; (ii) for each infinite word in L all its epsilon-approximations satisfy the specification. These verification results are strong in that they apply to all finite state Markov chains. Further, the study of the symbolic dynamics of Markov chains initiated here is of independent interest and can lead to other applications.

***Index Terms*—Model Checking, Probabilistic Computation, Approximation, Markov Processes.**

I. Introduction

Finite state Markov chains are a fundamental model of probabilistic dynamical systems. They are well-understood [13], [20] and their formal verification is well established [3]–[5], [8]–[10], [12], [14], [16], [17], [23]. In a majority of the verification related studies, the Markov chain is viewed a probabilistic transition system. The goal is to reason about the paths of the transition system using probabilistic temporal logics such as $PCTL$ [5], [10], [12].

An alternative view - which we follow here - is to view the state space of the chain to be the set of probability distributions over the nodes of the chain. The Markov chain transforms - in a linear fashion - a given probability distribution into a new one. Starting from a distribution μ one iteratively applies M to generate a trajectory consisting of a sequence of distributions. Given a set of initial distributions, one can study the properties of the set of trajectories generated by these distributions. The novel idea we explore in this setting is the *symbolic dynamics* of a Markov chain. We do so by discretizing the probability value space $[0, 1]$ into a *finite* set of intervals $\mathcal{I} = \{[0, p_1), [p_1, p_2), \ldots, [p_m, 1]\}$. A probability distribution μ of M over its set of nodes $\{1, 2, \ldots, n\}$ is then represented symbolically as a tuple of intervals $(d_1, d_2, \ldots, d_n)$ with d_i being the interval in which $\mu(i)$ falls. Such a tuple of intervals which symbolically represents at least one probability distribution is called a *discretized distribution*. In general a discretized distribution will represent an infinite set of concrete distributions.

A simple but crucial fact is that the set of discretized distributions, denoted $\mathcal{D}$, is a *finite* set. Consequently, each trajectory generated by an initial probability distribution will uniquely induce a sequence over the finite alphabet $\mathcal{D}$. Hence, given a (possibly infinite) set of initial distributions, the symbolic dynamics of M can be studied in terms of a language over the alphabet $\mathcal{D}$. Our focus here will be on infinite behaviors. Consequently, the main object of our study will be L_M, the ω-language over $\mathcal{D}$ induced by the set infinite trajectories generated by the set of initial distributions.

The main motivation for studying Markov chains in this fashion is to avoid the difficulties of numerically tracking sequences of probability distributions exactly. In many applications such as the probabilistic behavior of biochemical networks, queuing systems or sensor networks, exact estimates of the probability distributions (including the initial ones) may neither be feasible nor necessary. Further, not all the nodes may be relevant for the question at hand. In this case we can filter out such nodes by associating the "don't care" discretization $\{[0, 1]\}$ with each of them. This is a novel approach to dimension reduction and it can significantly reduce the practical complexity of analyzing high dimensional Markov chains. In our future work, we plan to apply this idea specifically to study the dynamics of biochemical networks.

To reason about the symbolic dynamics, we formulate a linear time temporal logic in which an atomic proposition will assert "the current probability of the node i lies in the interval d". The rest of the logic is obtained by closing under propositional connectives and the temporal modalities next and until in the usual way. We have chosen this simple logic in order to highlight the main ideas. As we point out in Section III this logic can be considerably strengthened and our techniques will easily extend to this strengthened version.

The key verification question is whether each sequence in L_M is a model of a specification φ. If L_M were to be a ω-regular language then standard model checking techniques will apply. Unfortunately determining whether L_M is ω-regular appears to be a difficult problem. Our current conjecture is - except in some special cases - L_M is *not* ω-regular. To sketch the nature of the problem, let us suppose that M is irreducible and aperiodic (the precise definition is given in Section V). This guarantees that M has a unique stationary distribution λ (i.e. $\lambda \cdot M = \lambda$). Further, every trajectory will converge to λ. However, if one or more components of λ coincide with the end-points of intervals in the discretization of $[0,1]$ then a trajectory may spiral towards λ while visiting the discretized distributions near λ in a non-periodic fashion. This is illustrated in fig. 1. Consequently the set of symbolic sequences generated by a set of initial distributions may fail to be ω-regular.

We bypass this basic difficulty by constructing approximate solutions to our verification problem. We fix an approximation factor $\epsilon > 0$. We then show that each symbolic trajectory in L_M can be split into a transient phase and a steady state phase. Further, if ξ_μ is the symbolic trajectory induced by the initial distribution μ, then in the steady state phase, ξ_μ will cycle through a set of final classes $\{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_{\theta-1}\}$ where each $\mathcal{F}_m$ is a set of discretized distributions. These final classes will be determined by M, the initial distribution μ and ϵ. Further, θ, the number of such classes will depend only on M. In addition, under a natural metric, all members of a final class

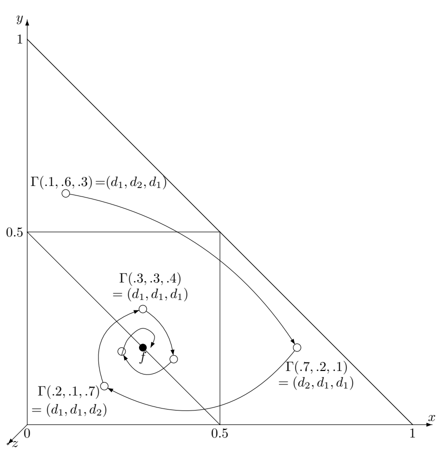


Fig. 1. A concrete and symbolic trajectory in the 3-dimensional space wrto. the discretization $\{d_1 = [0, 0.5), d_2 = [0.5, 1]\}$ projected onto the x, y plane, with values in x, y and z. (Γ is the map that associates discretized distributions to concrete ones with respect to the above discretization.)

will belong to a (discretized) ϵ-neighborhood. This leads to the notion of ϵ-approximation:

$\xi' \in \mathcal{D}^\omega$ is an ϵ-approximation of ξ_μ iff $\xi'(k) = \xi_\mu(k)$ for each k in the transient phase of ξ_μ. Moreover, $\xi'(k)$ and $\xi_\mu(k)$ are in the same final class and hence in an ϵ-neighborhood for each k in the steady state phase of ξ_μ.

This leads to two interesting notions of M ϵ-approximately meeting the specification φ. In stating these notions, we specify for convenience the initial set of concrete distributions as a discretized distribution D_{in}. In other words, μ is an initial distribution iff D_{in} is its symbolic representation.

1) (M, D_{in}) ϵ-approximately meets the specification φ from *below* - denoted $(M, D_{in}) \models_{\epsilon} \varphi$ - iff for every $\mu \in D_{in}$, *there exists* an ϵ-approximation of ξ_μ which is a model of φ.
2) (M, D_{in}) ϵ-approximately meets the specification φ from *above* -denoted $(M, D_{in}) \models^{\epsilon} \varphi$ - iff for every $\mu \in D_{in}$, *every* ϵ-approximation of ξ_μ is a model of φ.

Our main results are that given M, D_{in}, ϵ and φ, whether (M, D_{in}) ϵ-approximately satisfies φ from below (above) can be effectively determined. We note that $(M, D_{in}) \models^{\epsilon} \varphi$ implies that L_M itself meets the specification φ. On the other hand if it is not the case that $(M, D_{in}) \models_{\epsilon} \varphi$ then we can conclude that L_M does not meet the specification φ. The remaining case is when $(M, D_{in}) \models_{\epsilon} \varphi$ but it is not the case that $(M, D_{in}) \models^{\epsilon} \varphi$. Then, we can decide to accept that L_M meets the specification but only *ϵ-approximately*. In many applications, this will be adequate. If not, one can fix a smaller ϵ, say, $\frac{\epsilon}{2}$, and perform the two verification tasks again with minimal additional overhead.

We present only the main constructions and proof sketches here. All the details can be found in [1]. Further, we will often use basic results concerning Markov chains without an attribution. These results can be found in any standard text book; for instance [13], [20]. Finally, we do not address complexity issues in order to keep the focus on the main ideas. However, many of our constructions can be optimized and we plan to explore this important issue in the future.

Related work: Symbolic dynamics is a classical topic in the theory of dynamical systems [19]. Shift sequences is the key notion with shifts of finite type playing an important role in coding theory [18]. Here, instead, we focus on the symbolic dynamics from a verification standpoint.

Our discretization quotients the infinite set of probability distributions into a finite set of discretized distributions. In spirit, this resembles the regions based discretization in the theory of timed automata [2] which then leads to bisimulations of finite index. Similar constructions arise in the theory of hybrid automata [11] too. There are however two crucial differences. In our setting there are no resets involved and there is just one mode, namely the linear transform M, driving the dynamics. On the other hand, for timed automata and hybrid automata one obtains finite index bisimulations only in cases where the dynamics of the variables are decoupled from each other. In our setting this will be a deal breaker. Consequently our symbolic dynamics is delicately poised between "too coupled to analyze by using bisimulations of finite index", and "not expressive enough to lead to undecidability".

Viewing a Markov chain as a transform of probability distributions and carrying out formal verification of the resulting dynamics has been explored previously in [8], [14], [17]. In fact, the work reported in [8], [14] deals with MDPs (Markov Decison Processes) instead of Markov chains. However by considering the special case where the MDP accesses just one Markov chain we can compare our work with theirs. Firstly [8], [14], [17] consider only one initial distribution and hence just one trajectory needs to be analyzed. It is difficult to see how their results can be extended to handle multiple -and possibly infinitely many- initial distributions as we do. Secondly, they study only irreducible and aperiodic Markov chains. In contrast we consider the class of all Markov chains. Last but not least, they impose the drastic restriction that the unique fix point of the irreducible and aperiodic Markov chain is an *interior* point w.r.t. the discretization induced by the specification. In [8], a similar restriction is imposed in a slightly more general setting. Since the fix point is determined solely by the Markov chain and has nothing to do with the specification, this is not a natural restriction. We can also easily obtain an exact solution to our model checking problem by imposing such a restriction.

Returning to the two approaches to studying Markov chains, a natural question to ask is how they are related. It turns out that from a verification standpoint they are incomparable and complementary (see [7], [14]). Further, solutions to model checking problems in one approach (e.g. the decidability of $PCTL$ in the probabilistic transition system setting) will not translate into the other. Finally, intervals of probability distributions have been considered previously in a number of settings [15], [21], [24]. These studies focus on carrying out the usual numerical analysis within an envelope of upper and lower probability distributions and do not address issues related to symbolic dynamics. In contrast, we fix a discretization of $[0, 1]$ and develop a verification theory based on the induced symbolic dynamics.

Plan of the paper: In the next section, we define the notion of discretized distributions and the symbolic dynamics of Markov chains. In Section III, we introduce our temporal logic, illustrate its expressiveness and discuss how it can be extended. We then formulate our main results in Section IV. In Section V, we handle irreducible and aperiodic Markov chains; and in Section VI, irreducible but periodic chains. In the subsequent section general Markov chains are treated. In order to highlight the key technical issues, in these sections we consider just one initial concrete distribution. In Section VIII, we handle a set of initial concrete distributions. Future directions are discussed in the concluding section.

II. Discretized Distributions

Through the rest of the paper we fix a finite set of nodes $\mathcal{X} = \{1, 2, \ldots, n\}$ and let i, j range over $\mathcal{X}$. As usual a probability distribution over $\mathcal{X}$, is a map $\mu : \mathcal{X} \to [0, 1]$ such that $\sum_i \mu(i) = 1$. Henceforth we shall refer to such a μ as a distribution and sometimes as a concrete distribution. We let μ, μ' etc. to range over distributions. A Markov chain M over $\mathcal{X}$ will be represented as an $n \times n$ matrix with non-negative entries satisfying $\sum_j M(i, j) = 1$ for each i. Thus, if the system is currently at node i, then $M(i, j)$ is the probability of it being at j in the next time instant. We will say that M transforms μ into μ', if $\mu \cdot M = \mu'$.

We fix a partition of $[0, 1]$ into a finite set $\mathcal{I}$ of intervals and call it a *discretization*. We let d, d' etc. range over $\mathcal{I}$. Suppose $D : \mathcal{X} \to \mathcal{I}$. Then D is a *discretized distribution* iff there exists a concrete distribution μ such that $\mu(i) \in D(i)$ for every i. We denote by $\mathcal{D}$ the set of discretized distributions, and let D, D' etc. range over $\mathcal{D}$. A discretized distribution will sometimes be referred to as a *$\mathcal{D}$-distribution*. We often view D as an n-tuple $D = (d_1, d_2, \ldots, d_n) \in \mathcal{I}^n$ with $D(i) = d_i$.

Suppose $n = 3$ and $\mathcal{I} = \{[0, 0.2), [0.2, 0.4), [0.4, 0.7), [0.7, 1]\}$. Then $([0.2, 0.4), [0.2, 0.4), [0.4, 0.7))$ is a $\mathcal{D}$-distribution since for the concrete distribution $(0.25, 0.30, 0.45)$, we have $0.25, 0.30 \in [0.2, 0.4)$ while $0.45 \in [0.4, 0.7)$. On the other hand, neither $([0, 0.2), [0, 0.2), [0.2, 0.4))$ nor $([0.4, 0.7), [0.4, 0.7), [0.7, 1])$ are $\mathcal{D}$-distributions.

We have fixed a single discretization and applied it to each dimension to reduce notational clutter. As stated in the introduction, in applications, it will be useful to fix a different discretization $\mathcal{I}_i$ for each i. In this case one can set $\mathcal{I}_i = \{[0, 1]\}$ for each "don't care" node i. Our results will go through easily in such settings.

A concrete distribution μ can be abstracted as a $\mathcal{D}$-distribution D via the map Γ given by: $\Gamma(\mu) = D$ iff $\mu(i) \in D(i)$ for every i. Since $\mathcal{I}$ is a partition of $[0, 1]$ we are assured Γ is well-defined. Intuitively, we do not wish to distinguish between μ and μ' in case $\Gamma(\mu) = \Gamma(\mu')$. Note that $\mathcal{D}$ is a non-empty and *finite* set. By definition we also have that $\Gamma^{-1}(D)$ is a non-empty set of distributions for each D. Abusing notation -as we have been doing already- we will often view D as a set of concrete distributions and write $\mu \in D$ (or μ is in D etc.) instead of $\mu \in \Gamma^{-1}(D)$.

We focus on infinite behaviors. With suitable modifications, all our results can be specialized to finite behaviors. A *trajectory* of M is an infinite sequence of concrete distributions $\mu_0\mu_1 \ldots$ such that $\mu_l \cdot M = \mu_{l+1}$ for every $l \geq 0$. We let TRJ_M denote the set of trajectories of M (we will often drop the subscript M). As usual for $\rho \in TRJ$ with $\rho = \mu_0\mu_1 \ldots$, we shall view ρ as a map from $\{0, 1, \ldots\}$ into the set of distributions such that $\rho(l) = \mu_l$ for every l. We will follow a similar convention for members of $\mathcal{D}^\omega$, the set of infinite sequences over $\mathcal{D}$. Each trajectory induces an infinite sequence of $\mathcal{D}$-distributions via Γ. More precisely, we define $\Gamma^\omega : TRJ \to \mathcal{D}^\omega$ as $\Gamma^\omega(\rho) = \xi$ iff $\Gamma(\rho(\ell)) = \xi(\ell)$ for every ℓ. In what follows we will write Γ^ω as just Γ.

Given an initial set of concrete distributions, we wish to study the symbolic dynamics of M induced by this set of distributions. For convenience, we shall specify the set of initial distributions as a $\mathcal{D}$-distribution D_{in}. In general, D_{in}

will contain an infinite set of distributions. In the example introduced above, $([0.2, 0.4), [0.2, 0.4), [0.4, 0.7))$ is such a distribution. Our results will at once extend to sets of $\mathcal{D}$-distributions.

We now define $L_{M,D_{in}} = \{\xi \in \mathcal{D}^\omega \mid \exists \rho \in TRJ,\ \rho(0) \in D_{in},\ \Gamma(\rho) = \xi\}$. We view $L_{M,D_{in}}$ to be the symbolic dynamics of the system (M, D_{in}) and refer to its members as symbolic trajectories. From now on, we will write L_M instead of $L_{M,D_{in}}$ since D_{in} will be clear from the context.

Given (M, D_{in}), our goal is to specify and verify properties of L_M. If L_M were to be an ω-regular language then well-established techniques can be brought to bear. Unfortunately we do not know at present whether this is always the case. As explained in the introduction, we suspect that L_M is not ω-regular even for restricted Markov chains. In light of this, we shall approximately solve verification problems concerning L_M without placing any restrictions on M. However (as formalized in Prop. 2 below) our method can often yield exact verification results.

III. The Verification Problem

We formulate here the probabilistic linear time temporal logic $LTL_{\mathcal{I}}$. In the following sub-section we discuss how its expressive power can be extended. The set of atomic propositions is given by: $AP = \{\langle i, d\rangle \mid 1 \leq i \leq n, d \in \mathcal{I}\}$. The formulas of $LTL_{\mathcal{I}}$ are:

- Every atomic proposition is a formula.
- If φ and φ' are formulas then so are $\sim \varphi$ and $\varphi \vee \varphi'$.
- If φ is a formula then $O\varphi$ is also a formula.
- If φ and φ' are formulas then $\varphi U \varphi'$ is also a formula.

The atomic proposition $\langle i, d\rangle$ asserts that $D(i) = d$ where D is the current discretized distribution of M. This in turn means that if the current concrete distribution of M is μ then $\mu(i) \in d$. The propositional connectives such as $\wedge$, $\supset$ and $\equiv$ are derived in the usual way as also the unary modality $\Diamond$ via $\Diamond\varphi = tt U \varphi$ where tt is the constant formula that always evaluates to "true". This leads to $\Box\varphi = (\sim \Diamond \sim \varphi)$.

The semantics of the logic is captured by the satisfaction relation $\xi(l) \models \varphi$, where $\xi \in \mathcal{D}^\omega$, $l \geq 0$ and φ is a formula. This relation is defined inductively via:

- $\xi(l) \models \langle i, d\rangle$ iff $\xi(l)(i) = d$
- The connectives $\sim$ and $\vee$ are interpreted as usual.
- $\xi(l) \models O\varphi$ iff $\xi(l+1) \models \varphi$
- $\xi(l) \models \varphi U \varphi$ iff there exists $k \geq l$ such that $\xi(k) \models \varphi'$ and $\xi(l') \models \varphi$ for $l \leq l' < k$.

We say that ξ is a model of φ iff $\xi(0) \models \varphi$. As usual, L_φ is the set of models of φ. In what follows, for a distribution μ we let ρ_μ denote the trajectory in TRJ which satisfies: $\rho(0) = \mu$. We let $\xi_\mu = \Gamma(\rho_\mu)$ be the symbolic trajectory generated by μ.

We shall say that (M, D_{in}) meets the specification φ - and this is denoted $M, D_{in} \models \varphi$ - iff $\xi_\mu \in L_\varphi$ for every $\mu \in D_{in}$. Stated differently, $M, D_{in} \models \varphi$ iff $L_M \subseteq L_\varphi$. The model checking problem we wish to solve is: given a finite state Markov chain M, a discretization $\mathcal{I}$, an initial set of concrete distributions given as the discretized distribution D_{in} and a specification φ as an $LTL_{\mathcal{I}}$-formula, determine whether $M, D_{in} \models \varphi$.

We do not know at present whether this problem can be effectively solved, since it is not clear if and when L_M is a ω-regular language. Consequently we will solve this problem approximately. Before doing so we discuss what can be said in our logic and how its expressive power can be extended.

A. Expressiveness issues

Given a $\mathcal{D}$-distribution $D = (d_1, d_2, \ldots, d_n)$, we can assert that the current $\mathcal{D}$-distribution is D via $\bigwedge_i (i, d_i)$. We can assert D will appear infinitely often via $(\Box\Diamond \bigwedge_i (i, d_i))$. We can assert that the set of $\mathcal{D}$-distributions that appear infinitely often is from a given subset $\mathcal{D}'$ of $\mathcal{D}$ via $\Diamond\Box \bigvee_{D \in \mathcal{D}'} \langle D\rangle$ where $\langle D\rangle$ is an abbreviation for $\bigwedge_i (i, D(i))$. In fact one can easily strengthen this formula to assert that the set of $\mathcal{D}$-distributions that appear infinitely is exactly $\mathcal{D}'$. Suppose we classify members of $\mathcal{I}$ as representing "low" and "high" probabilities. For example, if $\mathcal{I}$ contains 10 intervals each of length 0.1, we can declare the first two intervals as "low" and the last two intervals as "high". In this case $\Box((i, d_9) \vee (i, d_{10}) \supset (j, d_1) \vee (j, d_2))$ will say that "whenever the probability of i is high, the probability of j will be low". We can gain considerable expressive power by letting an atomic proposition be a sentence taken from the first order theory of reals [22]. Specifically, we can define the predicate $dist(x_1, x_2, \ldots, x_n)$ to assert that $(x_1, x_2, \ldots, x_n)$ is a concrete distribution. We just have to say that each x_i is non-negative and that $\sum_i x_i = 1$. We can then say that the distribution $(x_1, x_2, \ldots, x_n)$ is in D via $x_i \in D(i)$ for each i. To be precise, $x_i \in D(i)$ is an abbreviation for $(l \leq x_i) \wedge (x_i < r)$ if $D(i) = [l, r)$ and similarly for the case $D(i) = [l, r]$. We can now form an arbitrary sentence ψ expressing a polynomial constraint over $\{x_1, x_2, \ldots, x_n\}$ saying that the distribution $dist(x_1, x_2, \ldots, x_n)$ satisfies ψ. Finally, we can assert that every concrete distribution in D satisfies ψ. For instance, we can say that $2(x_1 + x_2) < 3(x_3^2 + x_4)$ for every distribution (x_1, x_2, x_3, x_4) in the current discretized distribution D.

Next we can associate a value v_i (say, a rational number) with each node i denoting a physical quantity associated with the system. Then we can define the expected value $E(x_1, x_2, \ldots, x_n) = \sum_i x_i \cdot v_i$ and assert that "eventually, the expected value will always lie in the interval $(2, 3.5]$". In actual applications, we will have a *vector* of variables associated with M and these variables will denote the values of entities such as temperature, queue lengths, concentration levels of molecular species etc. Hence a rich set of quantitative properties can be captured by the atomic propositions and the time evolution of these quantities and their relationships can be captured in the extended logic. And all our results will go through for the extended logic as well.

A natural question is how logics interpreted over a sequence of probability distributions -such as $LTL_{\mathcal{I}}$ - are related to logics interpreted over the paths of a Markov chain -such as $PCTL$. As mentioned in the introduction, these two families of logics are *incomparable*. There is however much to explore here and we leave the exact delineation of what can be said in one framework but not the other to future research.

IV. The Main Results

To state our main results, we fix an approximation parameter $\epsilon > 0$. We expect ϵ to be a small fraction of the length of an interval (or the shortest interval) in $\mathcal{I}$. A crucial notion is that of a *discretized ϵ-neighborhood*. To capture this, we first define the distance Δ between two distributions μ and μ' as: $\Delta(\mu, \mu') = \sum_i |\mu(i) - \mu'(i)|$. The discretized ϵ-neighborhood of μ is denoted as $\mathcal{N}_\epsilon(\mu)$ and is the set of $\mathcal{D}$-distributions given by: $D \in \mathcal{N}_\epsilon(\mu)$ iff there exists $\mu' \in D$ such that $\Delta(\mu, \mu') \leq \epsilon$. Finally $\mathcal{F} \subseteq \mathcal{D}$ is a discretized ϵ-neighborhood iff there exists a distribution μ such that $\mathcal{N}_\epsilon(\mu) = \mathcal{F}$. For convenience, we will just say *ϵ-neighborhood* from now on.

Suppose $\mathcal{F}$ is an ϵ-neighborhood and $D_1, D_2 \in \mathcal{F}$. Then there exists μ_0 such that $\mathcal{N}_\epsilon(\mu_0) = \mathcal{F}$. Further there exist $\mu_1 \in D_1$ and $\mu_2 \in D_2$ such that $\Delta(\mu_1, \mu_0) \leq \epsilon$ and $\Delta(\mu_2, \mu_0) \leq \epsilon$. In this sense D_1 and D_2 will not be too far apart.

The key result concerning the symbolic dynamics is:

Proposition 1: Let M be a Markov chain, $\epsilon > 0$ and ξ_μ the symbolic trajectory generated by the distribution μ. Then, there exists (i) a positive integer θ that depends only on M (ii) a positive integer K^ϵ that depends only on M and ϵ and (iii) an ordered family of ϵ-neighborhoods $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ - called the *final classes* of μ - such that $\xi_\mu(k) \in \mathcal{F}_{\mu, k \mod \theta}$ for every $k > K^\epsilon$. Further, θ, K^ϵ and $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ are effectively computable.

According to this result, there will be a transient phase of length K^ϵ followed by a steady state phase in which ξ_μ will cycle through the ϵ-neighborhood families $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,2}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ forever. This leads to the notion of of an ϵ-approximation of a symbolic trajectory. Let μ be a distribution while θ, K^ϵ and $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ are as guaranteed by the above proposition. Then $\xi' \in \mathcal{D}^\omega$ is an *ϵ-approximation* of ξ_μ iff the following conditions hold:

- $\xi'(k) = \xi_\mu(k)$ for $0 \leq k \leq K^\epsilon$.
- For every $k > K^\epsilon$, $\xi'(k)$ belongs to $\mathcal{F}_{\mu, k \mod \theta}$.

Our approximate model checking problems can now be formalized:

Definition 1: Let M be a Markov chain, D_{in} an initial distribution, $\epsilon > 0$ an approximation factor and $\varphi \in LTL_{\mathcal{I}}$:

1) (M, D_{in}) ϵ-approximately meets the specification φ from *below*, denoted $M, D_{in} \models_\epsilon \varphi$, iff for every $\mu \in D_{in}$, it is the case that $\xi' \in L_\varphi$ for *some* ϵ-approximation ξ' of ξ_μ.
2) (M, D_{in}) ϵ-approximately meets the specification φ from *above*, denoted $M, D_{in} \models^\epsilon \varphi$, iff for every $\mu \in D_{in}$, it is the case that $\xi' \in L_\varphi$ for *every* ϵ-approximation ξ' of ξ_μ.

The two notions of approximate satisfaction yield valuable information about exact satisfaction as follows.

Proposition 2: Let M be a Markov Chain, $\epsilon > 0$ and φ be a property. Then

1) $(M, D_{in}) \models^\epsilon \varphi \implies (M, D_{in}) \models \varphi$, and
2) $(M, D_{in}) \not\models_\epsilon \varphi \implies (M, D_{in}) \not\models \varphi$.

The proof follows easily from the definitions and the observation that each ξ_μ is an ϵ-approximation of itself. Our main verification result is:

Theorem 1: Let M be a Markov chain, D_{in} an initial distribution, φ a specification and $\epsilon > 0$ an approximation factor. Then the questions whether $(M, D_{in}) \models_\epsilon \varphi$ and whether $(M, D_{in}) \models^\epsilon \varphi$ can both be effectively solved.

Here, we have fixed a discretization first and designed a temporal logic that is compatible with it by the choice of atomic propositions. Alternatively we could have started with a temporal logic which mentions point values of probabilities and used these probabilities as interval end points to fix a discretization; similar to the way regions and zones are derived in timed automata. We however feel that fixing a discretization independent of specifications and studying the resulting symbolic dynamics is a fruitful approach. Indeed, the discretization can be a crucial part of the modeling phase. One can then, if necessary, further refine the discretization in the verification phase.

We now turn to the proof of Theorem 1 by starting with irreducible and aperiodic Markov chains and considering increasingly complex classes. Further we shall first assume a single concrete initial distribution and then extend the results to a set of initial distributions.

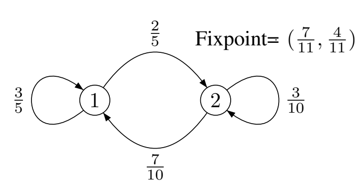
V. Irreducible and aperiodic Markov chains

Let M be a Markov chain over $\mathcal{X} = \{1, 2, \ldots, n\}$. The graph of M is the directed graph $G_M = (\mathcal{X}, E)$ with $(i, j) \in E$ iff $M(i, j) > 0$. We say that M is *irreducible* in case G_M is strongly connected. Assume M is irreducible. The period of the node i is the smallest integer m_i such that $M^{m_i}(i, i) > 0$. The period of M is denoted as θ_M and it is the greatest common divisor of $\{m_i\}_{i \in \mathcal{X}}$. The irreducible M is said to be *aperiodic* if $\theta_M = 1$. Otherwise it is *periodic*.

In what follows we shall abbreviate "irreducible and aperiodic" as just "aperiodic". Fig. 2 shows an example of an aperiodic Markov chain. Through the rest of this section, we will assume an aperiodic Markov chain over $\mathcal{X}$, a specification given as a $LTL_{\mathcal{I}}$-formula φ and an approximation factor $\epsilon > 0$. We also assume we are given a single initial distribution μ_0.

A. The determination of K^ϵ and the final classes

We set $\theta = \theta_M$. Since M is aperiodic, we have $\theta = 1$. To determine the final classes we start with the standard fact that the aperiodic chain M has a unique *stationary distribution* (fix point) λ. That is, $\lambda \cdot M = \lambda$. Further, every trajectory will converge to λ. One can effectively compute λ by solving the



$$M = \begin{vmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{7}{10} & \frac{3}{10} \end{vmatrix}$$

Fig. 2. An irreducible and aperiodic Markov chain M

linear system of equations $\mathbf{x} \cdot (M - I) = 0$ where I is the n-dimensional identity matrix. We then fix $\{\mathcal{F}_0 = \mathcal{N}_\epsilon(\lambda)\}$ to be the set of final classes.

Thus for the present case, there will be only one final class. If $\mathcal{F}_0$ itself is a singleton i.e. it consists of just one discretized distribution then it is easy to show that the original model checking problem can be solved exactly. In general however $\mathcal{F}_0$ will not be a singleton.

We next construct K^ϵ.

Lemma 1: There exists a positive integer K^ϵ such that for every μ and every $k > K^\epsilon$, we have $\Delta(\rho_\mu(k), \lambda) < \epsilon$.

Proof: One can effectively compute $\eta < 1$ such that for every μ we have $\Delta(\mu \cdot M^\ell, \lambda) < \eta \cdot \Delta(\mu, \lambda)$. The details can be found in [1]. Here, ℓ is the least integer -guaranteed to exist- such that $M^\ell(i,j) > 0$ for every i, j. It now follows that if $k > k' \cdot \ell$, then $\Delta(\mu \cdot M^k, \lambda) < \eta^{k'} \cdot \Delta(\mu, \lambda)$. In addition, by definition, $\Delta(\mu, \lambda) \leq 2$. Therefore, we can compute a sufficiently large k' and set $K^\epsilon = k' \cdot \ell$ so that for every μ, $\Delta(\mu \cdot M^k, \lambda) < \epsilon$ for every $k \geq K^\epsilon$. ■

It follows that $\rho_\mu(k) \in \mathcal{F}_0$ for every μ and every $k > K^\epsilon$, thus establishing Prop. 1 for the irreducible and aperiodic case.

B. Solutions to the approximate model checking problems

To determine whether $(M, \mu_0) \models_\epsilon \varphi$ we will construct a non-deterministic Büchi automaton $\mathcal{B}$ such that the language accepted by $\mathcal{B}$ is *non-empty* iff $(M, \mu_0) \models_\epsilon \varphi$. Since the emptiness problem for Büchi automata is decidable, we will have an effective solution to our model checking problem.

To start with, let $\Sigma = 2^{AP_\varphi}$ with AP_φ being the set of atomic propositions that appear in φ. We can interpret a formula of $LTL(\mathcal{I})$ over $\alpha \in \Sigma^\omega$ via: $\alpha(k) \models_\Sigma (i,d)$ iff $(i,d) \in \alpha(k)$ (ii) the propositional connectives are interpreted in the standard way. (iii) $\alpha(k) \models_\Sigma O(\varphi)$ iff $\alpha(k+1) \models_\Sigma \varphi$ (iv) $\alpha(k) \models_\Sigma \varphi_1 \boldsymbol{U} \varphi_2$ iff there exists $k' \geq k$ such that $\alpha(k') \models_\Sigma \varphi_2$ and $\alpha(k'') \models_\Sigma \varphi_1$ for $k \leq k'' < k'$. We say that α is a Σ-model of φ iff $\alpha(0) \models_\Sigma \varphi$. This leads to $\widehat{L}_\varphi = \{\alpha \mid \alpha, 0 \models \varphi\}$.

We can now construct the non-deterministic Büchi automaton $\mathcal{A} = (Q, Q_{in}, \Sigma, \longrightarrow, A)$ running over infinite sequences in Σ^ω such that the language accepted by $\mathcal{A}$ is exactly $\widehat{L}_\varphi$. This is a standard result [23] and we omit the details. We just wish to highlight that Q is the set of states, $Q_{in} \subseteq Q$ is the initial set of states, Σ is the alphabet, $\longrightarrow \subseteq Q \times \Sigma \times Q$ is the transition relation and $A \subseteq Q$ is the set of accepting states.

We next define $S = \{(k, \mu_0 \cdot M^k) \mid 0 \leq k \leq K^\epsilon\}$. The Büchi automaton $\mathcal{B} = (R, R_{in}, \Sigma, \Rightarrow, B)$ -which will also run over infinite sequences in Σ^ω- is defined as follows:

- $R = (S \cup \mathcal{F}_0) \times Q$ is the set of states.
- $R_{in} = \{(0, \mu_0)\} \times Q_{in}$ is the set of initial states.
- The transition relation $\Rightarrow$ is the least subset of $R \times \Sigma \times R$ satisfying the following conditions.
  First, suppose $((k,\mu), q)$ and $(k', \mu'), q'$ are in R and $Y \subseteq AP_\varphi$. Then $((k,\mu),q), Y, (k',\mu'), q')) \in \Rightarrow$ iff the following assertions hold:
  1) $k' = k + 1$ and $\mu \cdot M = \mu'$;
  2) if $(i,d) \in AP_\varphi$, then we have $\mu(i) \in d$ iff $(i,d) \in Y$;
  3) $(q, Y, q') \in \longrightarrow$.

  Next, suppose $((k,\mu), q))$ and (D, q') are in R with $D \in \mathcal{F}_0$. Let $Y \subseteq AP_\varphi$. Then $(((k,\mu),q), Y, (D, q')) \in \Rightarrow$ iff $k = K^\epsilon$ and $(i,d) \in Y$ iff $\mu(i) \in d$ for all $(i,d) \in AP_\varphi$. Furthermore, $(q, Y, q') \in \longrightarrow$.
  Finally, suppose (D, q) and (D', q') are in R and $Y \subseteq AP_\varphi$. Then $((D,q), Y, (D', q')) \in \Rightarrow$ iff for every $(i,d) \in AP_\varphi$, $D(i) = d$ iff $(i,d) \in Y$. Further, $(q, Y, q') \in \longrightarrow$.
- The set of final states is $B = \mathcal{F}_0 \times A$.

It is easy to show that $(M, \mu_0) \models_\epsilon \varphi$ iff the language accepted by $\mathcal{B}$ is non-empty. To determine whether $(M, \mu_0) \overset{\epsilon}{\models} \varphi$, we first construct the non-deterministic Büchi automaton $\mathcal{A}'$ such that the language accepted by $\mathcal{A}'$ is precisely $\hat{L}_{\sim\varphi}$. We then repeat the above construction using $\mathcal{A}'$ in place of $\mathcal{A}$ to construct $\mathcal{B}'$. Then one can show that $M, \mu_0 \overset{\epsilon}{\models} \varphi$ iff the language accepted by $\mathcal{B}'$ is *empty*. The details can be found in [1].

Notice that transitions of $\mathcal{B}$ and $\mathcal{A}$ check whether the current distribution μ satisfies $\mu(i) \in d$ or whether D satisfies $D(i) = d$. Since the first order theory of reals is decidable, we can also decide whether $\mu(i)$ or $D(i)$ satisfies a sentence ψ in this theory. Hence our decision procedures easily extend to the setting where atomic propositions consist of sentences in the first order theory of reals (as discussed in Section III).

VI. Irreducible Periodic Markov Chains

We now consider an irreducible periodic chain M *with period* $\theta_M \neq 1$. As before we will abbreviate “irreducible and periodic” as “periodic”. We set $\theta = \theta_M$. A standard fact is there exists a partitioning of $\mathcal{X}$ into θ equivalence classes $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_{\theta-1}$ such that in the graph of M, if there is an edge from i to j and $i \in \mathcal{X}_m$ then $j \in \mathcal{X}_{(m+1 \mod \theta)}$. In what follows we shall refer to each $\mathcal{X}_i$ as a component set. As before, we fix an initial distribution μ_0 and the approximation parameter $\epsilon > 0$. We let m, m' range over $\{0, 1, \ldots, \theta - 1\}$. An example of a periodic chain with period 3 is shown in fig. 3(a). In this chain, $\mathcal{X}_0 = \{1, 2\}$, $\mathcal{X}_1 = \{3\}$, $\mathcal{X}_2 = \{4\}$.

A key feature of M is that the probability masses of the component sets will cyclically shift through an application of M. To track this we will use the notion of a *weight vector* which is a map $\mathbf{w} : \{0, 1, \ldots, \theta - 1\} \to [0, 1]$ such that $\Sigma_m \mathbf{w}(m) = 1$. The distribution μ induces the weight vector $\mathbf{w}$ given by: $\mathbf{w}(m) = \Sigma_{i \in \mathcal{X}_m} \mu(i)$. Now suppose $\mu \cdot M = \mu'$

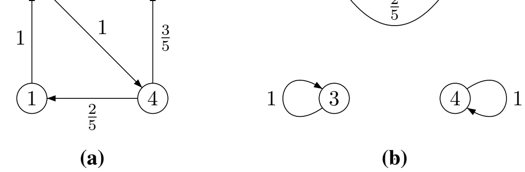


Fig. 3. An irreducible and periodic Markov chain and its class-decomposition

and $\mathbf{w}'$ is the weight vector induced by μ' .Then it is easy to check that $\mathbf{w}'(m+1 \mod \theta) = \mathbf{w}(m)$. As a consequence, if $\mathbf{w}''$ is the weight vector induced by $\mu \cdot M^\theta$ then $\mathbf{w} = \mathbf{w}''$.

A second key feature of M is that M^θ (i.e., M multiplied by itself θ times) restricted to $\mathcal{X}_m$ is an *aperiodic* chain for each m. We call these chains *component chains* and denote them as B_m for each m. We will obtain a global stationary distribution γ_0 of M^θ by weighting the unique local stationary distributions $\{\lambda_m\}$ of $\{B_m\}$ with the weight vector induced by μ_0. The ϵ-neighborhood of γ_0 will constitute a final class. Subsequently, the ϵ-neighborhoods of the global stationary distributions γ_m obtained via $\gamma_m = \gamma_0 \cdot M^m$ for each m will determine the θ final classes. We consider an example before formalizing this idea.

In fig. 3 we have shown the graphs of the three component chains B_0, B_1, B_2. The associated stationary distribution λ_0 of B_0 is given by $\lambda_0(1) = \frac{2}{5}$ and $\lambda_0(2) = \frac{3}{5}$. Clearly the stationary distributions of B_1 and B_2 are $\lambda_1(3) = 1$ and $\lambda_2(4) = 1$. The initial distribution $\mu_0 = (\frac{1}{5}, \frac{1}{10}, \frac{1}{2}, \frac{1}{5})$ induces the weight vector $\mathbf{w} = (\frac{3}{10}, \frac{1}{2}, \frac{1}{5})$. Hence γ_0 is given by $\gamma_0 = (\frac{6}{50}, \frac{9}{50}, \frac{1}{2}, \frac{1}{5})$. The other two stationary distributions from μ_0 are γ_1, γ_2, given by $\gamma_1 = \gamma_0 \cdot M$, $\gamma_2 = \gamma_0 \cdot M^2$.

A. The determination of K^ϵ and the final classes

As already fixed above, $\theta = \theta_M$ the period of M. In constructing the final classes and K^ϵ the basic observation is that the infinite sequence of distributions $(\mu_0 \cdot M^k)_{k\geq 0}$ can be analyzed in terms of the θ subsequences $(\mu_m \cdot M^{\theta \cdot k'})_{k' \geq 0}$ where $\mu_m = \mu_0 \cdot M^m$ for $0 \leq m \leq \theta - 1$. Actually one just needs to consider the first subsequence $(\mu_0 \cdot M^{\theta\cdot k'})_{k'\geq 0}$. The other subsequences can be simply obtained by applying M^m $(0 < m < \theta)$ to each element of this subsequence.

Before proceeding it will be convenient to introduce some additional notations using which we can analyze the global behavior of M^θ in terms of the local behaviors of the components $\{B_m\}$. Let $\mathbf{w}$ the weight vector induced by the distribution μ. Then $\downarrow^m (\mu) = \beta$ is the map $\beta : \mathcal{X}_m \to [0,1]$ given by: $\beta(i) = \frac{\mu(i)}{\mathbf{w}(m)}$ if $\mathbf{w}(m) \neq 0$ and $\beta(i) = 0$ if $\mathbf{w}(m) = 0$. It is easy to see that β is a distribution over $\mathcal{X}_m$ in case $\mathbf{w}(m) \neq 0$.

Next let $\{\beta_m\}$ be such that for each m, either β_m is a distribution over $\mathcal{X}_m$ or $\beta_m = 0_m$ (where $0_m(i) = 0$ for each $i \in \mathcal{X}_m$). Suppose $\mathbf{w}$ is a weight vector. Then $\uparrow^{\mathbf{w}} \{\beta_m\}$ is the map $\mu : \mathcal{X} \to [0,1]$ given by $\mu(i) = \mathbf{w}(m) \cdot \beta_m(i)$ if $i \in \mathcal{X}_m$. In the contexts in which we use this map, it will turn out that μ is a distribution over $\mathcal{X}$. In particular, if $\mathbf{w}$ is induced by μ then $\uparrow^{\mathbf{w}} \{\downarrow^m (\mu)\} = \mu$.

We are now prepared to define the final classes. Let λ_m be the unique stationary distribution of the component B_m for each m. Let $\mathbf{u}$ be the weight vector induced by μ_0. We now define $\gamma_0 = \uparrow^{\mathbf{u}} \{\lambda_m\}$. It is not difficult to check that γ_0 is a stationary distribution of M^θ. For each $m < \theta$, we let $\gamma_m = \gamma_0 \cdot M^m$. Finally, we define $\{\mathcal{F}_m = \mathcal{N}_\epsilon(\gamma_m)\}_{0 \leq m < \theta}$ as the set of final classes.

To determine K^ϵ we note that due to Lemma 1, for each component B_m we can effectively determine an integer $K_{\epsilon,m}$ such that for any distribution ν over $\mathcal{X}_m$, $\Delta(\lambda_m, \nu \cdot B_m^{K_{\epsilon,m}}) \leq \epsilon$. We now set $K^\epsilon = \theta \cdot \max_{m<\theta}\{K^{\epsilon,m}\}$.

To show that K^ϵ has the required properties we begin with:

Lemma 2: Suppose μ is a distribution and $\mathbf{w}$ is the weight vector induced by μ. Let $d \geq 0$ and $\mu \cdot M^{\theta \cdot d} = \mu'$. Then $\mu' = \uparrow^{\mathbf{w}} \{B_m^d(\downarrow^m (\mu))\}$.

The proof follows easily from the definitions.

Lemma 3: $\xi_{\mu_0(k)} \in \mathcal{F}_{(k \mod \theta)}$ for every $k > K^\epsilon$.

Proof: Let $\mu_0 \cdot M^k = \mu'$ with $k > K^\epsilon$. We shall show that $\Delta(\mu', \gamma_{(k \mod \theta)}) \leq K^\epsilon$. This will imply that $\xi_{\mu_0}(k) \in \mathcal{F}_{(k \mod \theta)}$ as required.

Let us consider first the case $k \mod \theta = 0$ with $k' = k \cdot \theta$. Let $\downarrow^m (\mu_0) = \beta_m$ for each m and assume $\mathbf{u}$ is the weight vector induced by μ_0. Set $\beta_m \cdot B_m^{k'} = \beta'_m$ for each m. For each m, in case $\mathbf{u}(m) = 0$, we have $\beta_m = 0_m$ and hence $\beta'_m = 0_m$. In case $\mathbf{u}(m) > 0$, by the choice of K^ϵ we are guaranteed $\Delta(\beta'_m, \lambda_m) \leq \epsilon$.

According to the previous lemma $\mu' = \uparrow^{\mathbf{u}} \{\beta'_m\}$. Hence $\Delta(\mu', \gamma_0) = \Sigma_m c_m$ where $c_m = \Sigma_{i\in\mathcal{X}_m} |\mathbf{u}(m) \cdot \beta'_m(i) - \mathbf{u}(m) \cdot \lambda_m| \leq \mathbf{u}(m) \cdot \epsilon$. Since $\Sigma_m \mathbf{u}(m) = 1$ we now have $\Delta(\mu', \gamma_0) \leq \epsilon$ as required. The other cases for $k = k'\theta + m$ for $0 < m < \theta$ follow easily from: $\Delta(\mu_0 \cdot M^k, \gamma_m) = \Delta(\mu_0 \cdot M^{k'\theta} \cdot M^m, \gamma_0 \cdot M^m) \leq \Delta(\mu_0 \cdot M^{k'\theta}, \gamma_0) \leq \epsilon$, due to the fact that $\Delta(\mu \cdot M, \mu' \cdot M) \leq \Delta(\mu, \mu')$ for any Markov chain M. ■

We have now established Prop. 1 for the class of periodic Markov chains.

B. Solutions to the approximate model checking problems

As before, we let $\Sigma = 2^{AP_\varphi}$ and first construct the non-deterministic Büchi automaton $\mathcal{A}$ such that the language accepted by $\mathcal{A}$ is exactly $\hat{L}_\varphi$ where $\hat{L}_\varphi$ is defined as in the previous section. The required Büchi automaton $\mathcal{B}$ can now be constructed along the lines followed in the previous section. Starting from μ_0, $\mathcal{B}$ will iteratively apply M and simulate $\mathcal{A}$ on the resulting $\mathcal{D}$-distributions. At the end of K^ϵ steps the resulting discretized distribution is guaranteed to be in $\mathcal{F}_0$. Starting from here, if the current $\mathcal{D}$-distribution is in $\mathcal{F}_m$ then the automaton will non-deterministically move to a $\mathcal{D}$-distribution in $\mathcal{F}_{m+1 \mod \theta}$ in the next step while continuing to simulate the automaton $\mathcal{A}$ on the resulting $\mathcal{D}$-distributions. The resulting run is accepted if $\mathcal{A}$ reports success. Otherwise it is rejected. We can then easily show;

Theorem 2: $M, \mu_0 \models_\epsilon \varphi$ iff the language accepted by $\mathcal{B}$ is non-empty.

To determine whether $M, \mu_0 \models^\epsilon \varphi$ we first construct the automaton $\mathcal{A}'$ which accepts $\hat{L}_{\sim\varphi}$. We then use it instead of the automaton $\mathcal{A}$ to construct a Büchi automaton $\mathcal{B}'$ such that $M, \mu_0 \models^\epsilon \varphi$ iff the language accepted by $\mathcal{B}'$ is *empty*. Again, all the details can be found in [1].

VII. Unrestricted Markov chains

Let M be a Markov chain with initial distribution μ_0, an approximation factor ϵ and a specification φ. We shall assume for convenience that G_M is connected and that for every node i either $\mu_0(i) > 0$ or there is a path $j_0 j_1 \ldots j_\ell$ in G_M such that $i = j_\ell$ and $\mu_0(j_0) > 0$. Let $\{SC_1, SC_2, \ldots, SC_r\}$ be the set of strongly connected components (SCCs) of G_M. The relation $\preceq$ over the SCCs is given by: $SC \preceq SC'$ iff there exists a node i in SC, a node j in SC' and a path from i to j in G_M. Clearly $\preceq$ is a partial ordering relation and the maximal elements under $\preceq$ will be called the bottom strongly connected components (BCCs). M restricted to each BCC will be aperiodic or periodic. If i belongs to a non-bottom SCC then it is a *transient* node. If a node is not transient then it is *recurrent*. An example of such a Markov chain is shown in fig. 4(a) and the poset of its SCCs is shown in fig. 4(b) (each SCC is represented by its set of nodes). Thus in fig. 4, nodes $\{1, 2, 3, 4\}$ are transient and $\{5, 6, 7, 8, 9\}$ are recurrent. The BCC $\{5, 6, 7\}$ is irreducible and periodic with period 3 while the BCC $\{8, 9\}$ is irreducible and aperiodic.

As M is iteratively applied to μ_0, the probability mass on the transient nodes will be transferred to the recurrent nodes. In the limit, *all* the probability mass will be distributed over the BCCs. We can then ignore the transient nodes and analyze a set of disjoint chains each of which will be aperiodic or periodic. This can be done easily using the results of the previous sections. However this will happen only in the limit while we must solve our model checking problems by running over sequences along which this transfer is taking place.

Our strategy will be to first compute the limit distribution of probability masses over the BCCs. This will enable us to compute a set of local final distributions induced by the BCCs. Then, as in the previous section, these local final distributions will be used to construct a set of global final distributions which in turn via their ϵ-neighborhoods will induce the final classes. We say here “final” instead of “stationary” to emphasize that these distributions will arise only in the limit.

The crucial next step is to compute two constants K_t and K_r such that by first iterating M on μ_0 for K_t times there will only be a negligible amount of probability mass left on the transient nodes. Then through a further K_r iterations we will be guaranteed to get ϵ-close to the desired global final distribution. With $K^\epsilon = K_t + K_r$ we then will be able to establish Prop.1 which in turn will lead to the solutions to the approximate model checking problems.

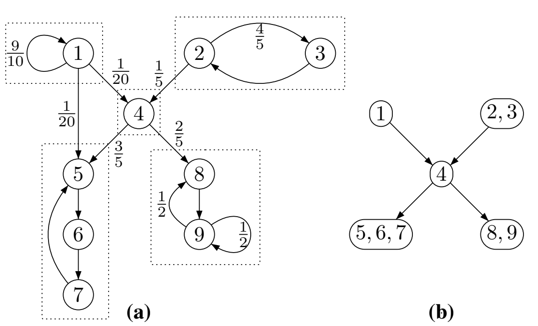


Fig. 4. A general Markov chain (unlabelled transitions have probability 1) and (the Hasse diagram of) its poset of strongly connected components

A. The determination of K^ϵ and the final classes

In what follows $\mathcal{X}_{trn}$ will denote the set of transient nodes and $\mathcal{X}_{rec}$ denote the set of recurrent states. We let $\{\mathcal{X}_1, \mathcal{X}_2, \dots \mathcal{X}_u\}$ be the node sets of the BCCs of M. Consequently, $\bigcup_{1\leq v\leq u} \mathcal{X}_v = \mathcal{X}_{rec}$. We let their periods be $\theta_1, \theta_2, \dots, \theta_u$, respectively. We next decompose each $\mathcal{X}_v$ - which in general will be periodic - into its set of aperiodic components $\mathcal{X}_{v,0}, \mathcal{X}_{v,1}, \dots, \mathcal{X}_{v,\theta_v-1}$ as done in the previous section. Obviously $\theta_v = 1$ and $\mathcal{X}_{v,0} = \mathcal{X}_v$ in case the v^{th} BCC is aperiodic. We define θ to be the lcm (least common multiple) of $\theta_1, \cdots, \theta_u$. As in the previous section, M^θ restricted to $\mathcal{X}_{v,m}$ will be denoted as $B_{v,m}$ and it will be an aperiodic chain. We shall refer to $\mathcal{X}_{v,m}$ as a component in what follows.

Let $\lambda_{v,m}$ be the unique stationary distribution of $B_{v,m}$ for each $v \in \{1, 2, \dots, u\}$ and each $m \in \{0, 1, \dots (\theta_v - 1)\}$. We next define the weight vector $\mathbf{w}$ which assigns the weight $\mathbf{w}(v, m)$ -denoting the probability mass transferred to $\mathcal{X}_{v,m}$ in the limit- to each (v, m). The key observation is that $\mathbf{w}$ can be obtained as the unique solution to a system of linear equations. Since the focus is on the components, we will need to work with M^θ rather than M. Accordingly, G^θ will denote the underlying graph of M^θ. We first define $T_{v,m}$ to be the subset of $\mathcal{X}$ given by: $i \in T_{v,m}$ iff $i \notin \mathcal{X}_{v,m}$ and there exists a directed path in G^θ from i to some node in $\mathcal{X}_{v,m}$. It is easy to see that $T_{v,m} \subseteq \mathcal{X}_{trn}$. There will be a variable x_i for each i in $T_{v,m}$ and the equation for x_i will be:

$x_i = \Sigma_{j\in T_{v,m}} M^\theta(i, j) \cdot x_j + \Sigma_{\ell\in\mathcal{X}_{v,m}} M^\theta(i, \ell)$.

Due to [6] (in particular Theorem 10.19) the above system of equations has a unique solution which can be computed effectively. Let $\mathbf{v} : T_{v,m} \to [0, 1]$ denote this solution. Again using the least fixpoint characterization of this solution provided in [6] we can in fact show that this solution satisfies $\mathbf{v}(i) = \lim_{k\to\infty} x_k(i)$, with $x_k(i) = \sum_{j\in\mathcal{X}_{v,m}} M^{k\cdot\theta}(i, j)$. We now define $\mathbf{w}(v, m) = \Sigma_{i\in T_{v,m}} \mu_0(i) \cdot \mathbf{v}(i) + \Sigma_{j\in\mathcal{X}_{v,m}} \mu_0(j)$. Using the expressions for $\mathbf{v}(i)$ as limits, it is easy to prove:

Lemma 4: $\forall v, m, k, \mathbf{w}(v, m) \geq \Sigma_{i\in\mathcal{X}_{v,m}} \mu_0 \cdot M^{k\cdot\theta}(i)$.

The global final distribution γ_0 is now given by $\gamma_0(i) = \mathbf{w}(v, m) \cdot \lambda_{v,m}(i)$ if $i \in \mathcal{X}_{v,m}$, and $\gamma_0(i) = 0$ if $i \in \mathcal{X}_{trn}$.

We let $\gamma_d = \gamma_0 \cdot M^d$ for $0 < d < \theta$. With $\mathcal{F}_d = \mathcal{N}_\epsilon(\gamma_d)$, this leads to $\{\mathcal{F}_d\}_{0\leq d\leq\theta-1}$ as the set of final classes.

The next task is to define K^ϵ. Let $i \in \mathcal{X}_{trn}$. Then there exists $k \leq |\mathcal{X}_{trn}|$ and $j \in \mathcal{X}_{rec}$ such that $M^k(i, j) > 0$. Next we note that if $i_1 \dots i_k$ is a path in G_M and $i_\ell \in \mathcal{X}_{rec}$ for some ℓ with $\ell < k$, then $i_v \in \mathcal{X}_{rec}$ for every v satisfying $\ell \leq v \leq k$. Consequently we can find a $p > 0$ such that for every $i \in \mathcal{X}_{trn}$, there exists $j \in \mathcal{X}_{rec}$ such that $M^{|\mathcal{X}_{trn}|}(i, j) \geq p$. This implies $\sum_{i\in\mathcal{X}_{trn}} (\mu_0 \cdot M^{|\mathcal{X}_{trn}|})(i) \leq (1 - p) \cdot \mu_0(i)$ for each $i \in \mathcal{X}_{trn}$. Since $p > 0$, we have $1 - p < 1$ and hence for any $\delta > 0$ there exists a computable k such that $(\mu_0 \cdot M^k)(i) < \delta$. Therefore we can fix K to be the least positive integer such that $\sum_{i\in\mathcal{X}_{trn}} (\mu_0 \cdot M^K)(i) \leq \frac{\epsilon}{4}$.

We now set $K_t = K \cdot \theta$ so that we have:

Lemma 5: for all $k \geq K_t$, $\sum_{i\in\mathcal{X}_{trn}} \mu_0 \cdot M^k(i) \leq \frac{\epsilon}{4}$.

Next we determine K_r by letting $K_{v,m}$ be the least integer such that for any distribution ν over $\mathcal{X}_{v,m}$, $|\nu \cdot B_{v,m}^{K_{v,m}}, \lambda_{v,m}| \leq$

$\frac{\epsilon}{2}$. Recall that $B_{v.m}$ is the restriction of M^θ to $\mathcal{X}_{v,m}$ and that $\lambda_{v,m}$ is the unique stationary distribution of the component $X_{v,m}$. We know from Lemma 3 that $K_{v,m}$ exists and is computable. We now set $K_r = \theta \cdot \max_{v,m}(K_{v,m})$. Finally, we fix K^ϵ as $K^\epsilon = K_t + K_r$.

To prove that K^ϵ has the required properties, we show that probability mass that has accumulated in $\mathcal{X}_{v,m}$ after K_t applications of M to μ_0 is close to $\mathbf{w}(v,m)$. More precisely, we have;

Lemma 6: $\Sigma_{v,m}|\mathbf{w}(v,m) - (\Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K^t}(j))| \leq \frac{\epsilon}{4}$.

Proof: By Lemma 4, $\Sigma_{v,m}|\mathbf{w}(v,m) - (\Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K^t}(j))| = \Sigma_{v,m}\mathbf{w}(v,m) - \Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K^t}(j)$. But $\Sigma_{v,m}\mathbf{w}(v,m) - \Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K^t}(j) = \Sigma_{v,m}\mathbf{w}_{v,m} - \Sigma_{v,m}\Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K_t}(j)$. Clearly $\Sigma_{v,m}\mathbf{w}(v,m) = 1$ and $\Sigma_{v,m}\Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K_t}(j) = \Sigma_{j\in\mathcal{X}_{rec}}\mu_0 M^{K_t}(j) = 1 - \Sigma_{j\in\mathcal{X}_{trn}}\mu_0 M^{K_t}(j) \geq 1 - \frac{\epsilon}{4}$ by Lemma 5. ∎

We now wish to show that by iterating further K_r times, starting from $\mu_0 \cdot M^{K_t}$, we will get ϵ-close to the final distribution γ_0. However it will be easier to work with the distribution μ' which is very much like $\mu_0 M^{K_t}$ but has no probability mass in $\mathcal{X}_{trn}$. We will then use μ' as a bridge to compare γ_0 with $\mu_0 \cdot M^{K^\epsilon}$. First define the weight vector β via $\beta(v,m) = \Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K_t}(j)$. We define μ' as:

- $\mu'(i) = 0$ for every $i \in \mathcal{X}_{trn}$.
- Suppose $j \in \mathcal{X}_{v,m}$. Then $\mu'(j) = \frac{\mathbf{w}(v,m)}{\beta(v,m)} \cdot \mu_0 M^{K_t}(j)$ if $\beta(v,m) \neq 0$. Else, $\mu'(j) = 0$.

Clearly, $\Sigma_{j\in\mathcal{X}_{v,m}}\mu'(j) = \mathbf{w}(v,m)$ for each (v,m).

The bridging role played by μ' can now be brought out:

Lemma 7: 1) $\Delta(\mu', \mu_0 \cdot M^{K_t}) \leq \frac{\epsilon}{2}$
2) $\Delta(\mu' \cdot M^{K_r}, \gamma_0) \leq \frac{\epsilon}{2}$

Proof: $\Delta(\mu', \mu_0 M^{K_t}) = \Sigma_{i\in\mathcal{X}}|\mu'(i) - \mu_0 M^{K_t}(i)| = \Sigma_{i\in\mathcal{X}_{trn}}|\mu'(i) - \mu_0 M^{K_t}(i)| + \Sigma_{j\in\mathcal{X}_{rec}}|\mu'(i) - \mu_0 M^{K_t}(i)|$. But then $\mu'(i) = 0$ for every $i \in \mathcal{X}_{trn}$ and $\Sigma_{i\in\mathcal{X}_{trn}}\mu_0 M^{K_t}(i) \leq \frac{\epsilon}{4}$ by Lemma 5.

We shall now show $\Sigma_{j\in\mathcal{X}_{rec}}|\mu'(j) - \mu_0 M^{K_t}(j)| \leq \frac{\epsilon}{4}$. Note $\Sigma_{j\in\mathcal{X}_{rec}}|\mu'(j) - \mu_0 M^{K_t}(j)| = \Sigma_{v,m}\Sigma_{j\in\mathcal{X}_{v,m}}|\mu'(j) - \mu_0 M^{K_t}(j)|$. From Lemma 4, it follows that $\mathbf{w}(v,m) \geq \beta(v,m)$ and hence $\mu'(j) \geq \mu_0 M^{K_t}(j)$ for every $j \in \mathcal{X}_{v,m}$. For all v,m, we have $\Sigma_{j\in\mathcal{X}_{v,m}}|\mu'(j) - \mu_0 M^{K_t}(j)| = \Sigma_{j\in\mathcal{X}_{v,m}}\mu'(j) - \mu_0 M^{K_t}(j) = \Sigma_{j\in\mathcal{X}_{v,m}}\mu'(j) - \Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K_t}(j) = \mathbf{w}(v,m) - \Sigma_{j\in\mathcal{X}_{v,m}}\mu_0 M^{K_t}(j)$. Summing over all v,m, we apply Lemma 5: $\Sigma_{j\in\mathcal{X}_{rec}}|\mu'(j) - \mu_0 M^{K_t}(j)| \leq \frac{\epsilon}{4}$.

The second part of the lemma follows the same line as the proof of Lemma 3. ∎

We can now obtain:

Lemma 8: $\Delta(\mu_0 \cdot M^{K^\epsilon}, \gamma_0) \leq \epsilon$

Proof: We have $\Delta(\mu_0 \cdot M^{K^\epsilon}, \gamma_0) = \Delta(\mu_0 \cdot M^{K_t} \cdot M^{K_r}, \gamma_0) \leq \Delta(\mu' \cdot M^{K_r}, \mu_0 \cdot M^{K_t} \cdot M^{K_r}) + \Delta(\mu' \cdot M^{K_r}, \gamma_0)$ using the triangle inequality. The second part is lower than $\frac{\epsilon}{2}$ due to the second statement of the lemma above. The first part is lower than $\frac{\epsilon}{2}$ due to the first part of the lemma above together with the fact that any Markov chain N satisfies $\Delta(y \cdot N, x \cdot N) \leq \Delta(y, x)$. ∎

Following the same line of reasoning as in the previous section, it is now easy to show that $\mu_0 \cdot M^k \in \mathcal{F}_{k \mod \theta}$ for every $k \geq K^\epsilon$. This establishes Prop. 1 for general Markov chains.

The construction of the required Büchi automata to solve the approximate model checking problems is then very similar to the ones in the previous sections and the details can be found in [1]. Thus we obtain:

Theorem 3: Given a specification φ we can effectively construct non-deterministic Büchi automata $\mathcal{B}, \mathcal{B}'$ such that:

1) $M, \mu_0 \models_{\epsilon} \varphi$ iff the language accepted by $\mathcal{B}$ is non-empty.
2) $M, \mu_0 \models^{\epsilon} \varphi$ iff the language accepted by $\mathcal{B}'$ is empty.

VIII. Multiple Initial Distributions

We now show how multiple initial distributions can be handled. Let M be a Markov chain. Assume that we are given a discretized distribution D_{in} as the set of initial concrete distributions.

Given $\mu \in D_{in}$, we now know how to compute - using μ in place of μ_0 - the final distribution γ_d to which $\mu_d \cdot M^{k\theta}$ will converge for $0 \leq d \leq \theta - 1$ and $\mu_d = \mu \cdot M^d$. However, we cannot handle members of D_{in} one at a time since there will be in general an infinite number of them. Hence we will group them into a finite number of equivalence classes as follows. Before proceeding it is worth noting that the choice of K^ϵ (as also that of K_r and K_t in the previous section) depended only on M and ϵ and not on the initial distribution μ_0. This is crucial for handling an infinite number of initial distributions.

Let μ be in D_{in} and let $\{\gamma^0_\mu, \ldots, \gamma^{\theta-1}_\mu\}$ be its associated set of final distributions. Then we will say that μ has the *ϵ-approximate behavior* $Bh = < D_1 D_2 \cdots D_{K^\epsilon}; \mathcal{D}_0, \ldots, \mathcal{D}_{\theta-1} >$ if $D_k = \xi_\mu(k)$ for $1 \leq k \leq K^\epsilon$, and $\mathcal{D}_d = \mathcal{N}_\epsilon(\gamma^d_\mu)$ for $0 \leq d < \theta$. Since $\mathcal{D}$ is a finite set there are only a finite number of ϵ-approximate behaviors.

Now suppose $\mu, \mu' \in D_{in}$ have the same ϵ-approximate behavior. Then it is easy to see that $(M,\mu) \models_{\epsilon} \varphi$ iff $(M,\mu') \models_{\epsilon} \varphi$. And $(M,\mu) \models^{\epsilon} \varphi$ iff $(M,\mu') \models^{\epsilon} \varphi$. This leads to the notion of $(M, Bh) \models_{\epsilon} \varphi$ which holds iff for some $\mu \in D_{in}$ whose ϵ-approximate behavior is Bh, we have $(M,\mu) \models_{\epsilon} \varphi$. Similarly $(M, Bh) \models^{\epsilon} \varphi$ holds iff for some $\mu \in D_{in}$ whose ϵ-approximate behavior is Bh, we have $(M,\mu) \models^{\epsilon} \varphi$. Clearly the algorithm of the previous section can be used to answer whether $(M, Bh) \models^{\epsilon} \varphi$ and whether $(M, Bh) \models_{\epsilon} \varphi$, for any ϵ-approximate behavior Bh. The issue now is which ϵ-approximate behaviors are witnessed (realized) by distributions in D_{in}.

To address this, we observe that D_{in} is a convex set of concrete distributions. In other words, if $\mu_1, \mu_2, \ldots, \mu_k \in D_{in}$ and $c_1, c_2, \ldots, c_k \in [0,1]$ with $\sum_l c_l = 1$ we are assured that $\mu = c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \ldots c_k \cdot \mu_k$ will be a distribution in D_{in}. Using the definition of a discretized distribution, we can easily find a finite set of *corner points* $CP = \{\kappa_1, \kappa_2, \ldots, \kappa_J\} \subseteq D_{in}$ such that for each $\mu \in D_{in}$ there exist $c_1, c_2, \ldots, c_J \in [0,1]$ such that $\sum_l c_l = 1$ and $\mu = c_1 \cdot \kappa^1 + c_2 \cdot \kappa^2 + \ldots + c_J \cdot \kappa^J$.

We wish to show that the final distributions induced by a distribution μ in D_{in} can be represented as the convex hull of the final distributions induced by the corner points. To make this precise, for each $\mu \in D_{in}$, let $\mathbf{w}_\mu$ be the weight vector induced by μ. Let γ_μ be the final distribution induced by μ as computed in the previous section, i.e. such that $\lim_{k\to\infty} \Delta(\mu \cdot M^{k\theta}, \gamma_\mu) = 0$.

Proposition 3: Let $\nu = \sum_{1\leq q\leq J} c_q \kappa_q$ with $c_q \in [0,1]$ and $\sum_{1\leq q\leq J} c_u = 1$. Then $\gamma_\nu = \sum_{1\leq q\leq J} c_q \gamma_{\kappa_q}$.

Proof: This follows from the linearity of M, that is, $(a\kappa + b\kappa')M = a\kappa M + b\kappa' M$. ■

By linearity, we also have a similar property for $\gamma_\mu^d = \gamma_\mu \cdot M^d$, $0 \leq d < \theta$. Hence, we only need to compute explicitly γ_κ for each corner point $\kappa \in CP$. Now, given a sequence $D_1 \cdots D_{K^\epsilon} \in \mathcal{D}$ and a θ-tuple of sets $(\mathcal{D}_0, \mathcal{D}_2, \ldots, \mathcal{D}_{\theta-1})$ with $\mathcal{D}_i \subseteq \mathcal{D}$ we can decide whether there exist $c_q \in [0,1]$ with $1 \leq u \leq J$ such that

- $\sum_{1\leq q\leq J} c_q = 1$,
- for all $k < K^\epsilon$, $\sum_{1\leq q\leq J} c_q M^k(\kappa_q) \in D_k$, and
- for all $0 \leq m \leq \theta - 1$, $\mathcal{N}_\epsilon(\sum_{1\leq q\leq J} c_q \gamma_{\kappa_q}^m) = \mathcal{D}_m$.

We can decide this using the first order theory of reals. Consequently we can compute the *finite* set of ϵ-approximate behaviors of M generated by the distributions in D_{in}. As noted earlier, for each ϵ-approximate behavior in this set, we can decide if it meets the specification φ and by taking the conjunction of all the outcomes we can decide whether $(M, D_{in}) \models_\epsilon \varphi$ and whether $(M, D_{in}) \models^\epsilon \varphi$.

IX. Conclusion

We have initiated here the study of the symbolic dynamics of finite state Markov chains obtained by discretizing the probability value space $[0,1]$ into a finite set of intervals. This leads to the notion of discretized distributions and symbolic trajectories. We have designed a simple temporal logic to reason about the symbolic dynamics and have considered two variants of an approximate model checking problem in this setting. Our main result is that both the variants are decidable.

In the present study we have used a discretized distribution to specify the initial set of distributions. An alternative approach would be to present it as the convex hull of a finite set of concrete distributions with rational component values. Naturally one can then also allow a finite union of such convex polytopes to specify the set of initial distributions. With some additional work our results can be easily extended to handle such initial distributions. Further, as pointed out at the end of Section V, we can also allow the atomic propositions to express polynomial constraints over the current distributions.

An interesting application to explore is the dynamics of biochemical networks modeled by the Chemical Master Equation [25]. We feel that our symbolic dynamics approach can bring considerable benefits in this setting. In particular, the errors incurred through the ϵ-approximation method will be entirely acceptable. Further applications can open up by extending our results to the setting of Markov decision processes (MDPs). As an orthogonal extension, one can also explore the discretization of transition probabilities. This will however considerably complicate the computation of the symbolic dynamics and hence will likely require the development of new techniques. Finally, as mentioned in the introduction, we have not paid close attention to complexity issues. We are however confident that geometric representations and linear algebraic techniques can considerably reduce the complexity of many of our constructions. We plan to address this issue as well in our future work.

Acknowledgment

Part of this work was done while Manindra Agrawal was visiting the IMS Workshop on Automata Theory at National University of Singapore. Support from the workshop and J C Bose Fellowship is gratefully acknowledged. We would also like to thank ANR project IMPRO and EU project DISC.

References

[1] M. Agrawal, S. Akshay, B. Genest and P. S. Thiagarajan. *Approximate verification of the symbolic dynamics of Markov chains*. Technical report available at *http://www.crans.org/~genest/AAGT12.pdf*.

[2] R. Alur and D. Dill. *A theory of timed automata*. TCS 126:183-235, 1994.

[3] C. Baier, E. Clarke, V. Hartonas-Garmhausen, M. Kwiatkowska and M. Ryan. *Symbolic model checking for probabilistic processes*. ICALP'97, LNCS 1256, pp. 430-440.

[4] C. Baier, B. Haverkort, H. Hermanns and J.-P. Katoen. *Model Checking meets Performance Evaluation*. ACM Performance Evaluation Review 32(4):10-15, 2005.

[5] C. Baier, H. Hermanns, J.-P. Katoen and V. Wolf. *Comparative Branching-Time Semantics for Markov Chains*. CONCUR'03, pp 482-497.

[6] C. Baier and J.-P. Katoen. *Principles of model checking*, MIT Press, 2008.

[7] D. Beauquier, A. Rabinovich and A. Slissenko. *A Logic of Probability with Decidable Model Checking*. CSL'02, pp. 306-321.

[8] R. Chadha, V. Korthikanti, M. Vishwanathan, G. Agha and Y. Kwon. *Model checking MDPs with a Unique Compact Invariant Set of Distributions*. QEST'11.

[9] V. Forejt, M. Kwiatkowska, G. Norman and D. Parker. *Automated Verification Techniques for Probabilistic Systems*. SFM'11, LNCS 6659, pp. 53-113.

[10] H. Hansson and B. Jonsson. *A logic for reasoning about time and reliability*. In Formal Aspects of Computing, 6(5):512-535, 1994.

[11] T. Henzinger, *The Theory of Hybrid Automata*. LICS'96, pp. 278-292.

[12] M. Huth and M. Kwiatkowska. *Quantitative Analysis and Model Checking*. LICS'97, pp. 111-122.

[13] J. G. Kemeny and J. L. Snell, *Finite Markov chains*, Princeton university press, 1960.

[14] V. Korthikanti, M. Vishwanathan, G. Agha and Y. Kwon, *Reasoning about MDPs as Transformers of probability distributions*. QEST'10, pp. 199-208.

[15] I. O. Kozine. *Interval-valued Markov chains*. Reliable Computing, 8, 2002 (97- 113).

[16] M. Kwiatkowska, G. Norman and D. Parker, *PRISM 4.0: Verification of Probabilistic Real-time Systems*. CAV'11, LNCS 6806, pp 585-591.

[17] Y. Kwon and G. Agha, *Linear Inequality LTL: A model checker for Discrete Time Markov chains*. ICFEM'04, pp. 194-208.

[18] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[19] M. Morse and G. A. Hedlund, *Symbolic Dynamics*, American Journal of Mathematics, 60:815-866, 1938.

[20] J. R. Norris. *Markov chains*, New York: Cambridge University Press, 1997. Cambridge series on statistical and probabilistic mathematics; No.2.

[21] D. Skulj. *Regular finite Markov chains with interval probabilities*. ISIPTA, 2007.

[22] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 1951.

[23] M. Y. Vardi. *Probabilistic linear-time model checking: An overview of the automata-theoretic approach*. ARTS'99, LNCS 1601, pp. 265-276.

[24] K. Weichselberger. *The theory of interval-probability as a unifying concept of uncertainty*. IJAR, 24(2-3):149-170, 2000.

[25] V. Wolf, R. Goel, M. Mateescu and T. Henzinger. *Solving the chemical master equation using sliding windows*. BMC Systems Biology 4:42, 2010.