ARTICLE IN PRESS
YINCO:4395

Information and Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Model checking Markov population models by stochastic approximations

Luca Bortolussi [a,b,*], Roberta Lanciani [c], Laura Nenzi [d,*]

[a] DMG, University of Trieste, Trieste, Italy
[b] MOSI, Department of Computer Science, Saarland University, Saarbrücken, Germany
[c] IMT, Lucca, Italy
[d] TU Wien, Austria

ARTICLE INFO

Article history:
Received 30 April 2017
Available online xxxx

Keywords:
Stochastic model checking
Fluid model checking
Stochastic approximation
Moment closure
Linear noise
Population models
Maximum entropy

ABSTRACT

Many complex systems can be described by population models, in which a pool of agents interacts and produces complex collective behaviours. We consider the problem of verifying formal properties of the underlying mathematical representation of these models, which is a Continuous Time Markov Chain, often with a huge state space. To circumvent the state space explosion, we rely on stochastic approximation techniques, which replace the large model by a simpler one, guaranteed to be probabilistically consistent. We show how to efficiently and accurately verify properties of random individual agents, specified by Continuous Stochastic Logic extended with Timed Automata (CSL-TA), and how to lift these specifications to the collective level, approximating the number of agents satisfying them using second or higher order stochastic approximation techniques.

1. Introduction

Many real-life examples of large complex systems, ranging from (natural) biochemical pathways to (artificial) computer networks, exhibit *collective behaviours*. These global dynamics are the result of intricate interactions between the large number of individual entities that comprise the populations of these systems. Understanding, predicting and controlling these emergent behaviours is becoming an increasingly important challenge for the scientists of the modern era. In particular, the development of an efficient and well-founded mathematical and computational modelling framework is essential to master the analysis of such complex collective systems.

In the Formal Methods community, powerful automatic verification techniques have been developed to validate the performance of a model of a system. In such *model checkers* [1], the model and a property of interest are given in input to an algorithm which verifies whether or not the requirement is satisfied by the representation of the system. As the dynamics of a collective system is intrinsically subject to noisy behaviours, especially when the population is not very large, the formal analysis and verification of a collective system have to rely on appropriate *Stochastic Model Checking* techniques. For instance, in [2], Continuous Stochastic Logic formulae are checked against models of the system expressed as Continuous Time Markov Chains (CTMC, [3]), which are a natural mathematical framework for population models. These approaches are based on an exhaustive exploration of the state space of the model, which limits their practical use, due to *state space*

* Corresponding authors.
E-mail addresses: luca@dmi.units.it (L. Bortolussi), roberta.lanciani@imtlucca.it (R. Lanciani), laura.nenzi@imtlucca.it (L. Nenzi).

https://doi.org/10.1016/j.ic.2018.09.004

Please cite this article in press as: L. Bortolussi et al., Model checking Markov population models by stochastic approximations, Inf. Comput. (2018), https://doi.org/10.1016/j.ic.2018.09.004

explosion: when the number of interacting agents in the population increases, the number of states of the underlying CTMC quickly reaches astronomical values. To deal with this problem, some of the most successful applications of Stochastic Model Checking to large population models are based on statistical analysis [4,27,5], which still remain costly from a computational point of view, because of the need or running simulation algorithms a large number of times.

In this work, we take a different approach, exploiting a powerful class of methods to accurately approximate the dynamics of the individuals and the population, that goes under the name of *Stochastic Approximations* [6].

Related work. Stochastic approximation methods have been successfully used in the computational biochemistry community [7,8] to approximate the noisy behaviour of collective systems by a simpler process whose behaviour can be extracted by solving a (numerically integrable) set of *Differential Equations* (DEs), resulting in a fast and easy way of obtaining an estimation of the dynamics of the model. Moreover, for almost all the techniques that we are going to consider in this work, the quality of the estimations improves as the number of agents in the system increases, keeping constant the computational cost and reaching exactness in the limit of an infinite population. In this way, such approximation methods actually take advantage of the large sizes of the collective systems, making them a fast, accurate and reliable approach to deal with the curse of the state space explosion. Among the many types of Stochastic Approximations present in the literature, we are going to exploit the *Fluid Approximation* (FA) [6,9], the *Central Limit Approximation* (CLA) [8,10], and the *System Size Expansion* (SSE) [11]. We are also going to use *Moment Closure* (MC) [11] combined with distribution reconstruction techniques based on the *Maximum Entropy* principle [12].

Stochastic Approximations entered into the model checking scene only recently. Pioneering work focussed on checking CSL properties [13,9,14] or deterministic automata specifications [15,16] for a single random individual in a population. Following this line of work, more complex individual properties had been considered, in particular rewards [17] and timed automata with one clock [18]. A similar approach for properties of individual agents of discrete time population models has beed considered in [19]. Another direction of integration of stochastic approximations and model checking is related to the so called local-to-global specifications [20], in which individual properties, specified by timed automata (with some restrictions), are lifted at the collective level by counting how many agents satisfy a local specification. This lifting is obtained by applying the CLA to approximate the satisfaction distribution of agents [20] or by moment closure to obtain bounds [15,16]. A simpler approach, focussing on expected values at the collective level, is [21]. Finally, stochastic approximation has been used also to approximate global reachability properties, either exploiting central limit results for hitting times [22], or by a clever discretisation of the Gaussian processes obtained by the CLA [23]. Recent related work on the global approximation of hitting times is [24], where authors use a Bayesian variational scheme to obtain a set of approximate ODEs for global hitting times, applied in [25] to CSL model checking.

Taking a broader perspective, the use of stochastic approximation can be seen as an abstraction procedure to simplify the analysis and verification of stochastic models. The literature in this sense is vast, ranging from abstraction refinement techniques [26] to statistical methods as in [27,5], and statistical abstraction [28]. Compared to numerical methods that still compute the Kolmogorov equation or are simulation-based, stochastic approximation provides a very fast alternative, with asymptotic guarantees, but lacking meaningful error bounds for finite populations, as the error bounds for stochastic approximation present in literature are too coarse, see e.g. the discussion in [29].

Contributions. In this paper, we start from the approach of [20] for the approximation of satisfaction probabilities of local-to-global properties, and extend it in several directions:

- We extend fluid model checking [9] to a subset of CSL-TA [30], a logic specifying temporal properties by means of Deterministic Timed Automata (DTA) with a single clock. We consider in particular DTAs in which the clock is never reset, and provide a model checking algorithm also for nested formulae, leveraging fluid approximation.
- We lift CSL-TA properties to the collective level, exploiting the central limit approximation, thus extending the approach of [20] to a more complex set of properties. We also remove some restrictions on the class of models considered with respect to those discussed in [20], and provide proofs of the relevant theorems.
- We extend both [20] and [9] by showing how to effectively use higher order approximations to correct for finite size effects. This requires to integrate within the model checking framework either moment closure or higher-order SSE [11], together with maximum entropy distribution reconstruction [12].

Throughout the paper, we make use of a simple but instructive running example of an epidemic model to illustrate the presented techniques.

Paper structure. The paper is organised as follows. In Section 2, we introduce the class of models we consider, and in Section 3, the property specification language. Section 4 contains an introduction on stochastic approximation techniques. Section 5 shows how to model check local properties described by CSL-TA, while Section 6 deals with local-to-global properties. Conclusions are drawn in Section 7. The appendix contains novel proofs.

2. Markov population models

In this section, we introduce a formalism to specify *Markovian Population Models*. These models consist of typically large collections of interacting components, or *agents*. Each component is a finite state machine, which can change internal state

Please cite this article in press as: L. Bortolussi et al., Model checking Markov population models by stochastic approximations, Inf. Comput. (2018), https://doi.org/10.1016/j.ic.2018.09.004