Probabilistic Model Checking for Continuous-Time Markov Chains via Sequential Bayesian Inference

Dimitrios Milios[1,2], Guido Sanguinetti[2,3], and David Schnoerr[2,4]

[1] Department of Data Science, EURECOM
[2] School of Informatics, University of Edinburgh
[3] SynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh
[4] Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London

Abstract. Probabilistic model checking for systems with large or unbounded state space is a challenging computational problem in formal modelling and its applications. Numerical algorithms require an explicit representation of the state space, while statistical approaches require a large number of samples to estimate the desired properties with high confidence. Here, we show how model checking of time-bounded path properties can be recast exactly as a Bayesian inference problem. In this novel formulation the problem can be efficiently approximated using techniques from machine learning. Our approach is inspired by a recent result in statistical physics which derived closed-form differential equations for the first-passage time distribution of stochastic processes. We show on a number of non-trivial case studies that our method achieves both high accuracy and significant computational gains compared to statistical model checking.

Keywords: Bayesian inference · model checking · moment closure.

1 Introduction

Probabilistic model checking of temporal logic formulae is a central problem in formal modelling, both from a theoretical and an applicative perspective [22, 1, 2, 4–6]. Classical algorithms based on matrix exponentiation and uniformisation are well-understood, and form the core routines of mature software tools such as PRISM [28], MRMC [26] and UPPAAL [7]. Nevertheless, the need to explicitly represent the state space makes their application to large systems problematic, or, indeed, theoretically impossible in the case of systems with unbounded state spaces, which appear frequently in biological applications.

Statistical model checking (SMC) approaches [37, 38] have emerged in recent years as a powerful alternative to exact techniques. Such methods provide a Monte Carlo estimate of the desired probability by repeatedly sampling trajectories from the model. SMC can also provide probabilistic guarantees on the

estimated probabilities, and, by choosing the number of simulations to be suitably large, one can reduce the uncertainty over the estimates arbitrarily.

While SMC offers a practical and flexible solution in many scenarios, its reliance on repeated simulation of the system makes it naturally computationally intensive. Although SMC can be trivially parallelized, the approach can still be computationally onerous for systems which are intrinsically expensive to simulate, such as systems with large agent counts or exhibiting stiff dynamics.

In this paper, we propose an alternative approach to solving the probabilistic model checking problem which draws on a recently proposed technique from statistical physics [33]. We show that the model checking problem is equivalent to a sequential Bayesian computation of the marginal likelihood of an auxiliary observation process. This marginal likelihood yields the desired time-bounded reachability probability, which is closely related to the eventually and globally temporal operators. We also expand the methodology to the case of the time-bounded until operator, thus covering a wide range of properties for temporal logics such as CSL [1, 2, 4–6]. The formulation of the model checking problem as a Bayesian inference method allows us to utilise efficient and accurate approximation methodologies from the machine learning community. In particular, we combine Assumed Density Filtering (ADF) [29, 31] with a moment-closure approximation scheme, which enables us to approximate the entire cumulative distribution function (CDF) of the *first* time that a time-bounded until property is satisfied by solving a small set of closed ordinary differential equations (ODEs) and low-dimensional integrals.

The rest of the paper is organised as follows. We discuss the related work in Section 2 and we provide some background material in Section 3. We then describe our new approach, highlighting both the links and differences to the recently proposed statistical physics method of [33] in Section 4. We consider four non-linear example systems of varying size and stiffness in Section 5.

2   Related Work

In recent years, the computational challenges of probabilistic model checking have motivated the development of approaches that rely on stochastic approximations as an alternative to both classical methods and SMC. In one of the earliest attempts, passage-time distributions were approximated by means of *fluid analysis* [24]. This framework was later extended to more general properties expressed as *stochastic probes* [15]. Fluid approximation has also been used to verify CSL properties for individual agents for large population models [9, 10]. In [11], a Linear Noise Approximation (LNA) was employed to verify not only local properties of individuals, but also global ones, which are given as the fraction of agents that satisfy a certain local specification. The verification of such local and global properties has been recently generalised for a wider class of stochastic approximations, including moment closure [13].

Regarding our work, one key difference with respect to these earlier approaches is that we consider global time-bounded until properties that char-

acterise the behaviour of the system at the population level. In that sense, our approach is mostly related to [12, 8], which rely on the LNA to approximate the probability of global reachability properties. In particular, the LNA is used to obtain a Gaussian approximation for the distribution of the hitting time to the absorbing set [12]. The methodology is different in [8], where it is shown that the LNA can be abstracted as a time-inhomogeneous discrete-time Markov chain which can be used to estimate time-bounded reachability properties. However, this method approximates the unconstrained process, and needs to subsequently resort to space and time discretisation to approximate the desired probability.

3   Background

A Continuous-Time Markov Chain (CTMC) is a Markovian (i.e. memoryless) stochastic process that takes values on a countable state space $\mathcal{S}$ and evolves in continuous time [18]. More formally:

Definition 1. *A stochastic process $\{X(t) : t \geq 0\}$ is a Continuous-Time Markov Chain if it satisfies the Markov property, i.e. for any $h \geq 0$:*

$$p(X_{t+h} = j \mid X_t = i, \{X_\tau : 0 \leq \tau \leq t\}) = p(X_{t+h} = j \mid X_t = i) \qquad (1)$$

A CTMC is fully characterised by its generator matrix Q, whose entries Q_{ij} denote the transition rate from state i to state j, for any $i, j \in \mathcal{S}$ [32]. The dynamics of a CTMC are defined by the *master equation*, which is a system of coupled ODEs that describe how the probability mass changes over time for all states. For a CTMC with generator matrix Q, the master equation will be:

$$\frac{dP(t)}{dt} = P(t)Q \qquad (2)$$

where $P(t)$ is the transition probability matrix at time t; the quantity $P_{ij}(t) = p(X_t = j \mid X_{t_0} = i)$ denotes the probability to transition from state i at time t_0 to state j at time $t \geq t_0$. The master equation is solved subject to initial conditions $P(0)$. Throughout this work, we shall consider CTMCs that admit a *population* structure, so that we can represent the state of a CTMC as a vector of non-negative integer-valued variables $\mathbf{x} = \{X_1, \ldots, X_N\}$, that represent population counts for N different interacting entities.

3.1   Moment Closure Approximation

For most systems, no analytic solutions to the master equation in (2) are known. If the state space $\mathcal{S}$ is finite, (2) constitutes a finite system of ordinary differential equations and can be solved by matrix exponentiation. For many systems of practical interest however, $\mathcal{S}$ is either infinite, or so large that the computational costs of matrix exponentiation become prohibitive.

Moment closure methods constitute an efficient class of approximation methods for certain types of master equations, namely if the elements Q_{ij} of the

generator matrix are polynomials in the state $\mathbf{x}$. This is for example the case for population CTMC of mass action type which are frequently used to model chemical reaction networks [20]. In this case, one can derive ordinary differential equations for the moments of the distribution of the process. Unless the Q_{ij} are all polynomials in $\mathbf{x}$ of order one or smaller, the equation for a moment of a certain order will depend on higher order moments, which means one has to deal with an infinite system of coupled equations. *Moment closure methods* close this infinite hierarchy of equations by truncating to a certain order. A popular class of moment closure methods does so by assuming $P(t)$ to have a certain parametric form [36]. This then allows to express all moments above a certain order in terms of lower order moments and thus to close the equations for these lower order moments.

In this paper, we utilise the so-called *normal moment closure* which approximates the solution of the master equation by a multi-variate normal distribution by setting all cumulants of order greater than two to zero [21, 34, 35]. This class of approximations was recently used within a formal modelling context in [19].

3.2   Probabilistic Model Checking

The problem of probabilistic model checking of CTMCs is defined in the literature as the verification of a CTMC against Continuous Stochastic Logic (CSL) [1, 2, 4–6]. A CSL expression is evaluated over the states of a CTMC. In the original specification [1], the syntax of a CSL formula is described by the grammar:

$$\phi ::= \mathtt{tt} \mid \alpha \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \mathcal{P}_{\bowtie p}(\Phi)$$

where ϕ is a state-formula, and Φ is a path-formula, i.e. it is evaluated over a random trajectory of the Markov chain. An atomic proposition α identifies a subset of the state space; in this paper, we consider atomic propositions to be linear inequalities on population variables. The probabilistic operator $\mathcal{P}_{\bowtie p}(\Phi)$ allows reasoning about the probabilities of a path-formula Φ:

$$\Phi ::= \mathbf{X}\phi \mid \phi_1 \, \mathbf{U} \, \phi_2 \mid \phi_1 \, \mathbf{U}^{[t_1,t_2]}\phi_2$$

$\mathcal{P}_{\bowtie p}(\Phi)$ asserts whether the probability that Φ is satisfied meets a certain bound expressed as $\bowtie p$, where $\bowtie \in \{\leq, \geq\}$ and $p \in [0,1]$. In order to evaluate the probabilistic operator, we need to calculate the satisfaction probability for a path-formula Φ, which involves one of three temporal operators: next $\mathbf{X}$, unbounded until $\mathbf{U}$, and time-bounded until $\mathbf{U}^{[t_1,t_2]}$.

For a finite CTMC, it is well-known that evaluating the probability of $\mathbf{X}\phi$ is reduced to matrix/vector multiplication, while evaluating the unbounded until $\phi_1 \, \mathbf{U} \, \phi_2$ requires solving a system of linear equations [4]. The time-bounded until operator can also be evaluated numerically via an iterative method that relies on uniformisation [4]. This process may have a prohibitive computational cost if the size of the state space is too large. For systems with unbounded state space, the only option to estimate the time-bounded until probabilities is by the means of stochastic simulation [37, 38], which also has a high computational cost.

Other temporal operators can be expressed as special cases of the until operator. For the time-bounded eventually operator we have: $\mathbf{F}^{[t_1,t_2]}\phi = \mathtt{tt}\,\mathbf{U}^{[t_1,t_2]}\phi$, while for the globally operator we have: $\mathbf{G}^{[t_1,t_2]}\phi = \neg\mathbf{F}^{[t_1,t_2]}\neg\phi$. The latter two operators formally describe the problem of time-bounded reachability.

4   Methodology

Given a property of the form $\Phi = \phi_1\,\mathbf{U}^{[0,t]}\phi_2$, our goal is to approximate the cumulative probability of reaching ϕ_2 at $\tau \leq t$, while satisfying ϕ_1 until then.

4.1   Time-bounded Reachability as Bayesian Inference

We first consider the problem of reachability, which is closely related to the eventually temporal operator $\mathbf{F}^{[0,t]}\phi$. If S_ϕ denotes the set of states that satisfy the formula ϕ, then we are interested in the probability that S_ϕ is reached for the first time; this quantity is also known in the literature as *first-passage time*.

Building upon [16, 17] Schnoerr et al [33] have recently formulated time-bounded reachability as a Bayesian inference problem. Using this formulation, they proposed a method where the entire distribution of first-passage times can be approximated by taking advantage of some well-established methodologies in the Bayesian inference and statistical physics literature. In the current section, we revise the approach of Schnoerr et al [33] for reachability, while in Section 4.2 we expand the method to the more general case of time-bounded until.

In the Markov chain literature [32], the states in the set S_ϕ are often called the *absorbing* states. Let $C = \mathcal{S} \setminus S_\phi$ denote the set of *non-absorbing* states. The cumulative probability for the system to reach an absorbing state at or before time t is equal to 1 minus the probability of the system having remained in C until t. Schnoerr et al's insight was to formulate this probability in terms of a Bayesian computation problem. Consider an auxiliary binary observation $c(t)$ process which evaluates to 1 whenever the system is in the non-absorbing set C and 0 otherwise. The pair $\{c(t), \mathbf{x}_t\}$ constitutes a *hidden Markov model* (HMM) in continuous time; the required cumulative probability would then correspond to the marginal likelihood of observing a string of all 1s as output of the HMM. Computing this marginal likelihood is a central and well studied problem in machine learning and statistics.

Even in this novel formulation, the problem is generally still intractable. To make progress, we first discretise the time interval $[0, t]$ into time points $\mathcal{T} = \{t_0 = 0, \ldots, t_N = t\}$ with spacing t/N. For the process $\mathbf{x}_{t_i}$ at time t_i being in C we thus have the observation model $p(C_{t_i}|\mathbf{x}_{t_i}) = 1$ if $\mathbf{x}_{t_i} \in C$ and zero otherwise. Note that $p(C_{t_i}|x_{t_i})$ is the distribution of the observation process $c(t)$, i.e. $c(t_i) \sim p(C_{t_i}|x_{t_i})$. The marginal likelihood $Z_{[0,t]}$ of having remained in C for all $t_i \in \mathcal{T}$ factorises as

$$Z_{[0,t]} = p(C_{t_0}) \prod_{i=1}^{N} p(C_{t_i}|C_{<t_i}) \tag{3}$$

where we introduced the notation $C_{<t_i} \equiv C_{t_{i-1},\dots,t_0}$. The factors of the rhs in (3) can be computed iteratively as follows. Let $\mathbf{x}_0$ be the initial condition of the process. Suppose that the system did not transition into the absorbing set until time t_{i-1} (that is, the process remained in C), and that the state distribution conditioned on this observations is $p(\mathbf{x}_{t_{i-1}}|C_{<t_i}, \mathbf{x}_0)$. We can solve the system forward in time up to time t_i to obtain the predictive distribution $p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)$, which will serve as a prior, and combine it with the likelihood term $p(C_{t_i}|\mathbf{x}_{t_i})$ that the process has remained in C at time t_i.

We can then define a posterior over the state space by applying the Bayes rule as follows:

$$p(\mathbf{x}_{t_i}|C_{\leq t_i}, \mathbf{x}_0) = \frac{p(C_{t_i}|\mathbf{x}_{t_i})\, p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)}{p(C_{t_i}|C_{<t_i}, \mathbf{x}_0)} \qquad (4)$$

The likelihood $p(C_{t_i}|\mathbf{x}_{t_i})$ represents the probability that the process does not leave C at time t_i. The prior denotes the state space probability considering that the process had remained in C for time $< t_i$. The posterior then will be the state space distribution after observing that the Markov process has remained in C at the current step.

The evidence $p(C_{t_i}|C_{<t_i}, \mathbf{x}_0)$ in (4) is a factor in the rhs of (3). It can be obtained by marginalising the joint probability $p(C_{t_i}, \mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)$ over $\mathbf{x}_{t_i}$:

$$p(C_{t_i}|C_{<t_i}, \mathbf{x}_0) = \int_{\mathcal{S}} p(C_{t_i}|\mathbf{x}_{t_i}) p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0) d\mathbf{x}_{t_i} \qquad (5)$$

The process described above is a Bayesian formulation for the introduction of absorbing states. By multiplying by the likelihood, we essentially remove the probability mass of transitioning to a state in S_ϕ; the remaining probability mass (the evidence) is simply the probability of remaining in C. Therefore, the probability of transitioning to S_ϕ for the first time at time t_i is the complement of the evidence:

$$p(S_{t_i}^\phi|C_{<t_i}, \mathbf{x}_0) = 1 - p(C_{t_i}|C_{<t_i}, \mathbf{x}_0) \qquad (6)$$

Thus, Equation (6) calculates the first-passage time probability for any $t_i \in \mathcal{T}$. Note that this approach neglects the possibility of the process leaving from and returning to region C within on time step. The time spacing thus needs to be chosen small enough for this to be a good approximation.

Schnoerr et al [33] further approximated the binary observation likelihood $p(C_{t_i}|\mathbf{x}_{t_i})$ by a soft, continuous loss function. This allowed them to take the continuum limit of vanishing time steps which in turn allows to approximate the evidence $p(C_{t_i}|C_{<t_i}, \mathbf{x}_0)$ by solving a set of ODEs. In this work, we keep the binary, discontinuous observation process and keep time discrete, which allows us to extend the framework from [33] to the time-bounded until operator.

4.2   The Time-bounded Until Operator

Consider the time-bounded property $\phi_1 \, \mathbf{U}^{[0,t]} \phi_2$ which will be satisfied if a state in S_{ϕ_2} is reached up to time t and the stochastic process has remained in S_{ϕ_1}

until then. Assuming that ϕ_1 is satisfied up to some $t_i \leq t$, there are three distinct possibilities regarding the satisfaction of the until property:

- it evaluated as false if we have $\mathbf{x}_{t_i} \notin S_{\phi_1}$ and $\mathbf{x}_{t_i} \notin S_{\phi_2}$ simultaneously,
- the property is evaluated as true if $\mathbf{x}_{t_i} \in S_{\phi_2}$,
- otherwise the satisfaction of the property is undetermined up to time t_i.

These possibilities correspond to three non-overlapping sets of states: $S_{\neg\phi_1 \wedge \neg\phi_2}$, S_{ϕ_2} and $S_{\phi_1} \setminus S_{\phi_2}$ accordingly, as seen in Figure 1.
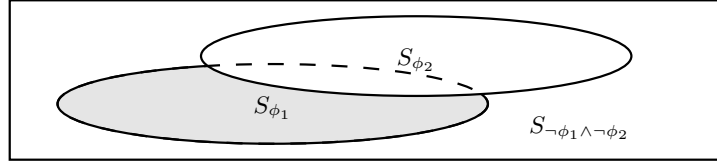
Fig. 1. The until formula $\phi_1 \, \mathbf{U}^{[0,t]} \phi_2$ is trivially satisfied for states in S_{ϕ_2}, while it is not satisfied for any state in $S_{\neg\phi_1 \wedge \neg\phi_2}$. For the rest of the states $C = S_{\phi_1} \setminus S_{\phi_2}$ (i.e. the grey area above) the property satisfaction is not determined. Assuming that the CTMC state has remained in C, we define a reachability problem to the union $S_{\phi_2} \cup S_{\neg\phi_1 \wedge \neg\phi_2}$. In contrast with the standard reachability problem, the probability of S_{ϕ_2} is of interest only, which is a subset of the absorbing states.

In order to calculate the satisfaction probabilities for any time $t_i \leq t$, we assume that the property has not been determined before t_i. That means that the Markov process has remained in the set $C = S_{\phi_1} \setminus S_{\phi_2}$, which is marked as the grey area in Figure 1. The Bayesian formulation of reachability discussed in Section 4.1 can be naturally applied to the problem of reaching the union $S_{\phi_2} \cup S_{\neg\phi_1 \wedge \neg\phi_2}$. The prior term $p(\mathbf{x}_{t_i} | C_{<t_i}, \mathbf{x}_0)$ denotes the state distribution given that the property remained undetermined before t_i. The likelihood term $p(C_{t_i} | \mathbf{x}_{t_i})$ indicates whether the Markov process has remained in the non-absorbing set $C = S_{\phi_1} \setminus S_{\phi_2}$ at t_i. Finally, the posterior given by (4) will be the state space distribution after observing that the property has remained undetermined at the last step.

In contrast with the reachability problem however, once the absorbing set is reached, we only know that the formula has been determined, but we do not know whether it has been evaluated as true or false. More specifically, the evidence $p(C_{t_i} | C_{<t_i}, \mathbf{x}_0)$ as given by Equation (5) represents the probability that the satisfaction has remained undetermined at time t_i. Although the negation of the evidence was sufficient to resolve the reachability probability as in Equation (6), we are now interested only in a subset of the absorbing states. At a particular time t_i we have to calculate the probability of reaching S_{ϕ_2} explicitly. This is given by the overlap mass of the prior process $p(\mathbf{x}_{t_i} | C_{<t_i}, \mathbf{x}_0)$ and probability of transitioning into S_{ϕ_2}. Given that $p(S_{t_i}^{\phi_2} | \mathbf{x}_{t_i}) = 1$, if $\mathbf{x}_{t_i} \in S_{\phi_2}$ at time t_i and zero otherwise, we have:

$$p(S_{t_i}^{\phi_2} | C_{<t_i}, \mathbf{x}_0) = \int_{\mathcal{S}} p(S_{t_i}^{\phi_2} | \mathbf{x}_{t_i}) p(\mathbf{x}_{t_i} | C_{<t_i}, \mathbf{x}_0) d\mathbf{x}_{t_i} \tag{7}$$

which is the probability of transitioning to S_{ϕ_2} at time t_i, while remaining in C until then. The effect of $p(S_{t_i}^{\phi_2}|\mathbf{x}_{t_i})$ is essentially a truncation of the state space; the first-passage probability at t_i is simplified as follows:

$$p(S_{t_i}^{\phi_2}|C_{<t_i}, \mathbf{x}_0) = \int_{\mathbf{x}_{t_i} \in S_{\phi_2}} p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)d\mathbf{x}_{t_i} \tag{8}$$

Considering a Gaussian approximation for $p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)$, as we discuss in the next section, and given that the state formula ϕ_2 is a conjunction of linear inequalities, Equation (8) can be easily calculated by numerical routines.

The Bayesian formulation that we introduce has essentially the same effect as the traditional probabilistic model checking methods [5]. The probability of the until operator is usually evaluated by first introducing the set of absorbing states $S_{\phi_2} \cup S_{\neg\phi_1 \wedge \neg\phi_2}$, and then calculating the probability of reaching the set S_{ϕ_2}, which is a subset of the absorbing states. The advantage of the formulation presented here is that it allows us to leverage well-established machine learning methodologies, as we see in the section that follows.

4.3   Gaussian Approximation via Assumed Density Filtering

The Bayesian formulation as described so far does not involve any approximations apart from time discretisation. In fact for a discrete-state system, both the prior and the likelihood terms (i.e. $p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0)$ and $p(C_{t_i}|\mathbf{x}_{t_i})$ equivalently) will be discrete distributions in (4). Therefore, quantities such as the evidence in (5) and the probability of reaching S_{ϕ_2} in Equation (8) can be calculated exactly, as the integrals reduce to summations. However, if the size of the state space is too large or unbounded, this process can be computationally prohibitive. The formulation presented above allows us to derive an efficient approximation method that relies on approximating the discrete process by a continuous one.

We adopt a moment closure approximation scheme where all cumulants of order three or larger are set to zero, which corresponds to approximating the single-time distribution of the process by a Gaussian distribution. As described in Section 3.1, the moment closure method results in a system of ODEs that describe the evolution of the expected values and the covariances of the population variables in a given CTMC. At any time t_i, the state distribution is approximated by a Gaussian with mean μ_{t_i} and covariance Σ_{t_i}:

$$p(\mathbf{x}_{t_i}|C_{<t_i}, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t_i}; \mu_{t_i}, \Sigma_{t_i})$$

The evidence is the probability mass of non-absorbing states; i.e. it is observed that the process has remained within C. Since C is identified by linear inequalities on the population variables, both the evidence in Equation (5) and the probability mass in the target set in (8) can be estimated by numerically solving the integrals. There are many software routines readily available to calculate the CDF of multivariate Gaussian distributions by numerical means.

Nevertheless, the posterior in Equation (4) is not Gaussian and we introduce a Gaussian approximation. It is proven that ADF minimises the KL divergence

between the true posterior and the approximating distribution, subject to the constraint that the approximating distribution is Gaussian [29, 31]. Considering the prior $\mathcal{N}(\mathbf{x}_{t_i}; \mu_{t_i}, \Sigma_{t_i})$, the ADF updates [17] will be:

$$\tilde{\mu}_{t_i} = \mu_{t_i} + \Sigma_{t_i} \partial_{\mu_{t_i}} \log Z_{t_i} \tag{9}$$

$$\tilde{\Sigma}_{t_i} = \Sigma_{t_i} + \Sigma_{t_i} \partial^2_{\mu^2_{t_i}} \log Z_{t_i} \Sigma_{t_i} \tag{10}$$

where the evidence $Z_{t_i} = p(C_{t_i} | C_{<t_i}, \mathbf{x}_0)$ is equal to the mass of the truncated Gaussian that corresponds to the non-absorbing states C at time t_i. The dimensionality of the Gaussians is equal to the number of distinct populations in the system; this is generally small, meaning that computations of truncated Gaussian integrals can be carried out efficiently. A detailed exposition can be found in the archive version of the paper [30].

4.4 Algorithm

Algorithm 1 is an instantiation of model checking via sequential Bayesian inference (MC-SBI). The algorithm evaluates the probability that a property $\Phi = \phi_1 \mathbf{U}^{[0,t]} \phi_2$ is satisfied for a sequence of time points $\mathcal{T} = \{t_0 = 0, t_1, \ldots, t_N = t\}$, thus approximating the CDF of the *first* time that Φ is satisfied.

Algorithm 1 Model Checking via Sequential Bayesian Inference

Require: CTMC with initial state $\mathbf{x}_0$, property $\Phi = \phi_1 \mathbf{U}^{[0,t]} \phi_2$, time sequence $\mathcal{T} = \{0, t_1, \ldots, t\}$
Ensure: Probabilities $\{\pi_0, \ldots, \pi_N\}$ that approximate the CDF of the time that Φ is satisfied
1: Define $C = S_{\phi_1} \setminus S_{\phi_2}$, where the satisfaction of Φ is not determined
2: Set the initial prior: $p(\mathbf{x}_{t_0} | C_{<t_0}, \mathbf{x}_0) \leftarrow \mathcal{N}(\mathbf{x}_{t_0}; \mu_{t_0}, \Sigma_{t_0})$
3: Initialise the probability that Φ is not determined: $p(C_{<t_0}, \mathbf{x}_0) \leftarrow 1$
4: **for** $i \leftarrow 0$ **to** N **do**
5:     Calculate the probability that Φ is satisfied for first time at t_i:

$$\pi_i \leftarrow p(C_{<t_i}, \mathbf{x}_0) \times \int_{\mathbf{x}_{t_i} \in S_{\phi_2}} p(\mathbf{x}_{t_i} | C_{<t_i}, \mathbf{x}_0) d\mathbf{x}_{t_i}$$

6:     Calculate the evidence $p(C_{t_i} | C_{<t_i}, \mathbf{x}_0)$ according to Equation (5)
7:     Calculate the probability that Φ is not determined in the next step:

$$p(C_{<t_{i+1}}, \mathbf{x}_0) \leftarrow p(C_{<t_{i+1}}, \mathbf{x}_0) \times p(C_{t_i} | C_{<t_i}, \mathbf{x}_0)$$

8:     Calculate the posterior mean $\tilde{\mu}_{t_i}$ and covariance $\tilde{\Sigma}_{t_i}$ according to (9) and (10) respectively
9:     Considering $\tilde{\mu}_{t_i}$ and $\tilde{\Sigma}_{t_i}$ as initial conditions,
       use moment closure ODEs to obtain: $\mu_{t_{i+1}}$ and $\Sigma_{t_{i+1}}$
10:    Set the prior of the next step:

$$p(\mathbf{x}_{t_{i+1}} | C_{<t_{i+1}}, \mathbf{x}_0) \leftarrow \mathcal{N}(\mathbf{x}_{t_{i+1}}; \mu_{t_{i+1}}, \Sigma_{t_{i+1}})$$

11: **end for**

In the beginning of each iteration at line 5, we calculate the probability π_i that Φ is satisfied at t_i. In lines 6–8, we calculate the posterior state distribution, assuming that Φ has not been determined at the current step. Finally, the state distribution is propagated by the moment closure ODEs; the new state probabilities $p(\mathbf{x}_{t_{i+1}} | C_{<t_{i+1}}, \mathbf{x}_0)$ will serve as the prior in the next iteration.

It is useful at this stage to pause and consider the differences from the first-passage time algorithm proposed in [33]: both papers share the same insight that reachability properties can be computed via Bayesian inference. However, the resulting algorithms are quite different. The crucial technical difficulty when considering formulae involving an until operator is the need to evaluate the probability of transitioning into the region identified by the second formula S_{ϕ_2}. It is unclear how to incorporate such a computation within the continuous-time differential equations approach of [33], which dictates the choice of pursuing a time discretisation approach here. The time discretisation however brings the additional benefit that we can evaluate *exactly* the moments of the Bayesian update in step 8 of Algorithm 1, thus removing one of the sources of error in [33] (at a modest computational cost, as the solution of ODEs is generally faster than the iterative approach proposed here).

5   Examples

In this section, we demonstrate the potential of our approach on a number of examples. More specifically, we report for each example the calculated CDF for the time that a formula $\Phi = \phi_1 \, \mathbf{U}^{[0,t]} \phi_2$ is first satisfied. Additionally for each until property, we also report the CDF of the first-passage time to the absorbing set; this corresponds to the eventually formula $\mathbf{F}^{[0,t]}(\phi_2 \vee \neg\phi_1 \wedge \neg\phi_2)$, following the discussion of Section 4.2.

As a baseline reference, we use the PRISM Model Checker [28], which is a well-established tool in the literature. For a time-bounded until property Φ, PRISM is capable of estimating its satisfaction probability by considering the following variation of the probabilistic operator $\mathcal{P}_{=?}(\Phi)$. The result of $\mathcal{P}_{=?}(\Phi)$ denotes the probability that Φ has been satisfied at any $\tau \leq t$, thus it can be directly compared to our approach. In particular, PRISM offers numerical verification of time-bounded until properties that relies on the uniformisation method [4]. We make use of numerical verification when possible, but for more complex models we resort to SMC.

5.1   An epidemiology model

We consider a SIR model, whose state is described by three variables that represent the number of susceptible (X_S), infected (X_I), and recovered (X_R) individuals in a population of fixed size. The dynamics are described by the reactions:

$S + I \xrightarrow{k_i} I + I$, with rate function $k_i X_S X_I$;
$I \xrightarrow{k_r} R$, with rate function $k_r X_I$;
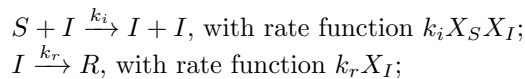
Considering initial state $[X_S = 40, X_I = 10, X_R = 0]$, the reachable state space as reported by PRISM involves 1271 states and 2451 transitions, which is a number small enough to allow the use of numerical verification.

We consider two properties: the first property states whether the infected population remains under a certain threshold until the extinction of the epidemic:

$$\varphi_1 = X_I < 30 \, \mathbf{U}^{[0,t_1]} X_I = 0 \tag{11}$$
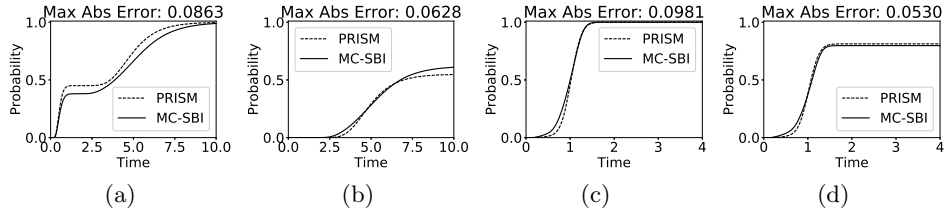
Fig. 2. First-passage time results for the SIR model: (a) the CDF of first-passage times into the absorbing states for φ_1, (b) CDFs of first-passage times for the until formula φ_1, (c) the CDF of first-passage times into the absorbing states for φ_2, (d) CDFs of first-passage times for the until formula φ_2.

where $t_1 = 10$. Also, we consider a property that involves more than one species:

$$\varphi_2 = X_S > 1 \, \mathbf{U}^{[0,t_2]} X_I < X_R \tag{12}$$

where $t_2 = 4$. The random variables $[X_S, X_I, X_R, X_I - X_R]$ follow a joint Gaussian distribution, which is compatible with the assumptions of our approach.

We have used Algorithm 1 to approximate the CDF of the time that φ_1 and φ_2 are first satisfied on a sequence $\mathcal{T}$ of 200 time-points. We have also used the hybrid engine of PRISM in order to produce accurate estimates of the satisfaction probabilities of φ_1 and φ_2, for $t_1 \in [0, 10]$ an $t_2 \in [0, 4]$ respectively.

The calculated CDFs for φ_1 are summarised in Figure 2b, while in Figure 2a we report the CDFs of the first-passage time into its absorbing set. Similarly, the CDFs for φ_2 are reported in 2d, and the CDFs of the corresponding absorbing set can be found in Figure 2c. In both cases the CDFs calculated by our approach (MC-SBI) are close to the numerical solutions of PRISM.

5.2   LacZ - A model of prokaryotic gene expression

We consider the model of LacZ protein synthesis in E. coli that first appeared in [27] and has been used as a model checking benchmark [14]. The model consists of 12 species and 11 reactions; its full specification can be found in [30]. We are interested in three variables: $X_{Ribosome}$ for the population of ribosomes, $X_{TrRbsLacZ}$ which represents the population of translated sequences, and X_{LacZ} representing the molecules of protein produced. The following property:

$$\varphi_3 = (X_{Ribosome} > 0 \wedge X_{TrRbsLacZ} < 200) \, \mathbf{U}^{[0,500]} X_{LacZ} > 150 \tag{13}$$

monitors whether both $X_{Ribosome}$ and $X_{TrRbsLacZ}$ satisfy certain conditions until the LacZ protein produced reaches a specified threshold (i.e. $X_{LacZ} > 150$). A randomly sampled trajectory can be seen in Figure 3a.

We have attempted to explore the reachable state space of the model using the hybrid engine of PRISM; that involved more than 26 trillions of states and 217 trillions of transitions. It is fair to state that numerical methodologies can
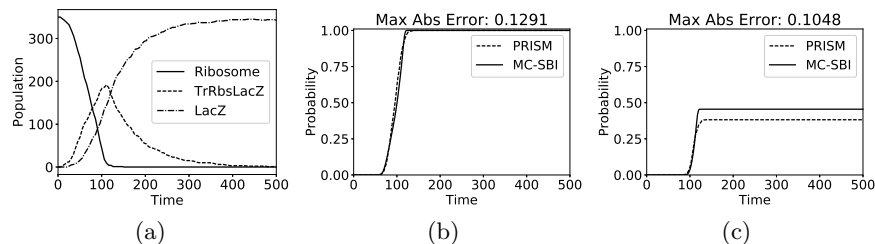
Fig. 3. First-passage time results for the LacZ model: (a) sample trajectory, (b) the CDF of first-passage times into the absorbing states for the property φ_3, (c) CDFs of first-passage times for the until formula φ_3.

be ruled out for this example. Thus we compare our approach with SMC as implemented in PRISM, where 1000 samples were used; the confidence interval for the results that follow is $\pm$ 0.039, based on 99.0% confidence level.

Figure 3 summarises the calculated first-passage time CDFs evaluated on a sequence of 200 time-points. In Figure 3b we see that the moment closure method resulted in a particularly accurate approximation of the first-passage time distribution for the absorbing states. Regarding the distribution of φ_3, the results of MC-SBI and PRISM's SMC seem to be in agreement (Figure 3c); however that our method overestimates the final probability of satisfying φ_3.

5.3   A stiff viral model

Stiffness is a computational issue in many chemical reaction systems which arises when some reactions occur much more frequently than others. This group of fast reactions dominates the computational time, and thus renders simulation particularly expensive. As an example of a stiff system, we consider the model of viral infection in [23]. The model state is described by four variables: the population of the viral template X_T, the viral genome X_G, the viral structural protein X_S, and X_V that captures the number of viruses produced. For the initial state we set $X_T = 10$, and the rest of the variables to zero. The reactions that determine the dynamics can be found in [30]. The model state space is unbounded, therefore we resort to the SMC capabilities of PRISM to evaluate our approach. The SMC used 1000 samples, resulting in confidence interval $\pm$ 0.038, based on 99.0% confidence level.

Figure 4a depicts a random trajectory that shows the evolution of the viral genome X_G and the virus population X_V over time. We see that X_G slowly increases until it apparently reaches a steady-state and fluctuates around the value 200, while X_V continues to increase at a non-constant rate. In this example, we shall monitor whether the viral genome remains under the value of 200 until the virus population reaches a certain threshold:

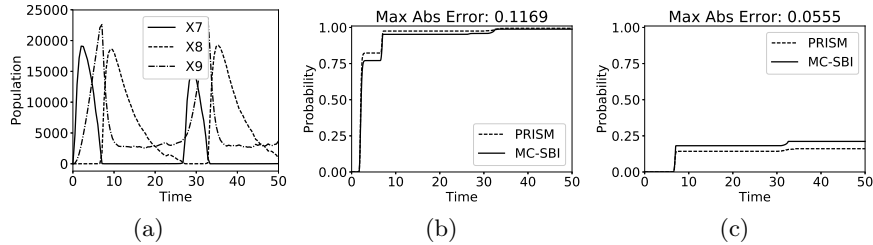$$\varphi_4 = X_G < 200\,\mathbf{U}^{[0,200]}X_V > 500 \tag{14}$$
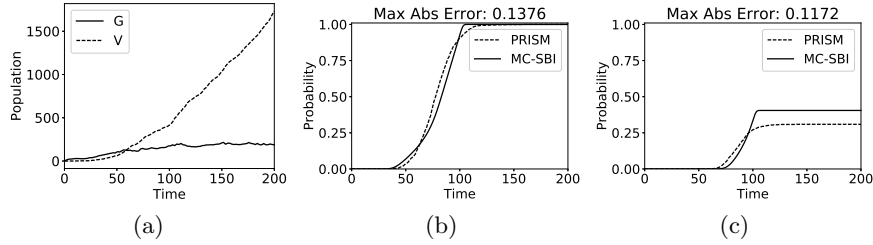
Fig. 4. First-passage time results for the Viral model: (a) sample trajectory, (b) the CDF of first-passage times into the absorbing states for the property φ_4, (c) CDFs of first-passage times for the until formula φ_4.

The results of Figure 4 show that our method did not capture the CDFs as well as in the previous two examples. However, considering that our method is four orders of magnitude faster than SMC (cf. Table 1), it still gives a reasonably good approximation, particularly in the case of the eventually value. Again, we have considered a sequence of length 200.

5.4 A genetic oscillator

Finally, we consider the model of a genetic oscillator in [3] consisting of 9 species and 16 reactions. The original model is defined in terms of concentrations; in order to convert the specification in terms of molecular populations, we consider a volume $V = 1/6.022 \times 10^{-22}$. The full model description can be found in [30]. We consider an initial state where $X_1 = 10$, $X_3 = 10$ and the rest of the variables are equal to 1. As we can see in the random trajectory in Figure 5a, we have a system too large to apply traditional model checking methods.

Fig. 5. First-passage time results for the genetic oscillator model: (a) sample trajectory, (b) the CDF of first-passage times into the absorbing states for the property φ_5, (c) CDFs of first-passage times for the until formula φ_5.

We focus on variables X_7 and X_9; the following property monitors whether X_7 remains under 19000 until X_9 exceeds the value of 24000:

$$\varphi_5 = X_7 < 19000 \, \mathbf{U}^{[0,50]} X_9 > 24000 \tag{15}$$

In this example, we evaluated the CDF on a sequence $\mathcal{T}$ of length 2000. We compare against the SMC algorithm in PRISM using 1000 samples, which resulted in confidence interval $\pm$ 0.030, based on 99.0% confidence level. Figure 5 shows a good approximation of the rather unusual first-passage time CDFs for both the absorbing states (Figure 5b) and the φ_5 property (Figure 5c).

5.5   A note on the execution times

Table 1 summarises the execution times for our method (MC-SBI) and statistical model checking (SMC). We have used numerical verification as implemented in PRISM for the SIR model only. For the other examples, the state space is too large to use an explicit representation, and we report the simulation running times only. The numerical approach is much faster when this is applicable. However, the computational savings for MC-SBI are obvious for the more complicated examples, in particular the viral model and the genetic oscillator.

We have used StochDynTools [25] to derive the moment closure approximations automatically. The CDFs have been evaluated on a sequence of 200 points for all models except for the genetic oscillator, where 2000 points were used.

Table 1. Execution times in seconds for model checking via sequential Bayesian inference (MC-SBI) and model checking in PRISM (10^4 samples were used for SMC).

Model	MC-SBI	PRISM (Numerical)	PRISM (SMC)
SIR	8 sec	$\sim$ 1 sec	$\sim$ 1 sec
LacZ	38 sec	N/A	46 sec
Viral	8 sec	N/A	24875 sec
Genetic Oscillator	87 sec	N/A	20707 sec

6   Conclusions

We have presented a novel approach to the classical model checking problem based on a reformulation as a sequential Bayesian inference problem. This reformulation is exact up to time-discretisation errors; it was originally suggested in [33] for reachability problems, and was extended in the present work to general CSL formulae including time-bounded Until operators. Apart from its conceptual appeal, this reformulation is important because it enables us to obtain an approximate solution using efficient and highly accurate tools from machine learning. Our method leverages a class of analytical approximations to CTMCs known as moment closures, which enable an efficient computation of the process marginal statistics.

We have shown on a number of diverse case studies that our method achieves excellent accuracy with significantly reduced computational costs compared to SMC. Nevertheless, our algorithm requires some approximations to the underlying stochastic process. The first approximation is the adoption of a time discretisation; this is a controllable approximation and can be rendered arbitrarily precise by reducing the time step (at a computational cost that grows linearly

with the number of steps). The second approximation consists in propagating forward the first two moments of the process via moment closure and ADF. Approximation quality in this case is system dependent. Several studies have examined the problem of convergence of moment closure approximations [34, 35], however, to the best of our knowledge, error bounds for such approximations are an open problem in the mathematics of stochastic processes. Despite such issues, we believe that the reformulation of model checking problems in terms of Bayesian inference has the potential to open the door to a new class of approximate algorithms to attack this classic problem in computer science.

References

1. A. Aziz, K. Sanwal, V. Singhal, and R. Brayton. Verifying continuous time Markov chains. In *Proc. of CAV*, pages 269–276. Springer, 1996.
2. A. Aziz, K. Sanwal, V. Singhal, and R. Brayton. Model-checking continuous-time Markov chains. *ACM Trans. Comput. Logic*, 1(1):162–170, 2000.
3. P. Azunre, C. Gomez-Uribe, and G. Verghese. Mass fluctuation kinetics: analysis and computation of equilibria and local dynamics. *IET Syst. Biol.*, 5(6):325–335, 2011.
4. C. Baier, B. Haverkort, H. Hermanns, and J.-P. Katoen. Model checking continuous-time Markov chains by transient analysis. In *Proc. of CAV*, pages 358–372. Springer, 2000.
5. C. Baier, B. Haverkort, H. Hermanns, and J.-P. Katoen. *Automated Performance and Dependability Evaluation Using Model Checking*, pages 261–289. Springer Berlin Heidelberg, 2002.
6. C. Baier, B. Haverkort, H. Hermanns, and J. P. Katoen. Model-checking algorithms for continuous-time Markov chains. *IEEE Trans. Softw. Eng.*, 29(6):524–541, 2003.
7. G. Behrmann, A. David, K. G. Larsen, J. Hakansson, P. Petterson, W. Yi, and M. Hendriks. UPPAAL 4.0. In *Proc. of QEST 2006*, pages 125–126. IEEE Computer Society, 2006.
8. L. Bortolussi, L. Cardelli, M. Kwiatkowska, and L. Laurenti. Approximation of probabilistic reachability for chemical reaction networks using the linear noise approximation. In *Proc. of QEST*, pages 72–88. Springer, 2016.
9. L. Bortolussi and J. Hillston. Fluid model checking. In *Proc. of CONCUR*, pages 333–347. Springer, 2012.
10. L. Bortolussi and J. Hillston. Model checking single agent behaviours by fluid approximation. *Inf. Comput.*, 242(C):183–226, 2015.
11. L. Bortolussi and R. Lanciani. Model checking markov population models by central limit approximation. In *Proc. of QEST*, pages 123–138. Springer, 2013.
12. L. Bortolussi and R. Lanciani. Stochastic approximation of global reachability probabilities of markov population models. In *Proc. of Computer Performance Engineering*, pages 224–239. Springer, 2014.
13. L. Bortolussi, R. Lanciani, and L. Nenzi. Model checking markov population models by stochastic approximations. *CoRR*, abs/1711.03826, 2017.
14. L. Bortolussi, D. Milios, and G. Sanguinetti. Smoothed model checking for uncertain continuous time Markov chains. *Inform. Comput.*, 247:235–253, 2016.
15. J. T. Bradley, R. A. Hayden, and A. Clark. Performance specification and evaluation with unified stochastic probes and fluid analysis. *IEEE Trans. Softw. Eng.*, 39:97–118, 2013.

16. B. Cseke, M. Opper, and G. Sanguinetti. Approximate inference in latent gaussian-markov models from continuous time observations. In *Proc. of NIPS*, pages 971–979. Curran Associates Inc., 2013.
17. B. Cseke, D. Schnoerr, M. Opper, and G. Sanguinetti. Expectation propagation for continuous-time stochastic processes. *J. Phys. A-Math. Theor.*, 49(49):494002, 2016.
18. R. Durrett. *Essentials of Stochastic Processes*. Springer, 2012.
19. C. Feng, J. Hillston, and V. Galpin. Automatic moment-closure approximation of spatially distributed collective adaptive systems. *ACM Trans. Model. Comput. Simul.*, 26(4):26:1–26:22, 2016.
20. C. W. Gardiner. *Handbook of Stochastic Methods*, volume 3. Springer Berlin, 1985.
21. L. A. Goodman. Population growth of the sexes. *Biometrics*, 9(2):212–225, 1953.
22. H. Hansson and B. Jonsson. A logic for reasoning about time and reliability. *Form. Asp. Comput.*, 6(5):512–535, 1994.
23. E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys*, 117(15):6959, 2002.
24. R. A. Hayden, A. Stefanek, and J. T. Bradley. Fluid computation of passage-time distributions in large markov models. In *Proc. of QAPL*, volume 413, pages 106–141, 2012.
25. J. P. Hespanha. `StochDynTools` — a MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions.
26. J.-P. Katoen, M. Khattri, and I. S. Zapreevt. A Markov reward model checker. In *Proc. of QEST*, pages 243–244. IEEE Computer Society, 2005.
27. A. M. Kierzek. STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics*, 18(3):470–481, 2002.
28. M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In *Proc. of CAV*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.
29. P. S. Maybeck. *Stochastic Models, Estimation, and Control*. Academic Press, 1982.
30. D. Milios, G. Sanguinetti, and D. Schnoerr. Probabilistic model checking for continuous-time markov chains via sequential bayesian inference. *CoRR ArXiv*, abs/1711.01863v2, 2018.
31. T. P. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, MIT, 2001.
32. J. R. Norris. *Markov chains*. Cambridge University Press, 1997.
33. D. Schnoerr, B. Cseke, R. Grima, and G. Sanguinetti. Efficient low-order approximation of first-passage time distributions. *Phys. Rev. Lett.*, 119:210601.
34. D. Schnoerr, G. Sanguinetti, and R. Grima. Validity conditions for moment closure approximations in stochastic chemical kinetics. *J. Chem. Phys*, 141(8):08B616_1, 2014.
35. D. Schnoerr, G. Sanguinetti, and R. Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys*, 143(18):11B610_1, 2015.
36. D. Schnoerr, G. Sanguinetti, and R. Grima. Approximation and inference methods for stochastic biochemical kinetics – a tutorial review. *J. Phys. A*, 50(9):093001, 2017.
37. H. L. Younes and R. G. Simmons. Statistical probabilistic model checking with a focus on time-bounded properties. *Inf. Comput.*, 204(9):1368–1409, 2006.
38. P. Zuliani, A. Platzer, and E. M. Clarke. Bayesian statistical model checking with application to Simulink/Stateflow verification. In *Proc. of HSCC*, pages 243–252. ACM, 2010.