Product Form Approximation of Transient Probabilities in Stochastic Reaction Networks

Alessio Angius and András Horváth

Department of Computer Science, University of Torino, Torino, Italy
angius@di.unito.it, horvath@di.unito.it

Abstract

Most Markov chains that describe networks of stochastic reactions have a huge state space. This makes exact analysis infeasible and hence the only viable approach, apart from simulation, is approximation. In this paper we derive a product form approximation for the transient probabilities of such Markov chains. The approximation can be interpreted as a set of interacting time inhomogeneous Markov chains with one chain for every reactant of the system. Consequently, the computational complexity grows only linearly in the number of reactants and the approximation can be carried out for Markov chains with huge state spaces. Several numerical examples are presented to illustrate the approach.

Keywords: stochastic reaction network; Markov chain; transient analysis; product form approximation.

1 Introduction

In [9,10], Gillespie provided a stochastic description of the evolution of a general chemical reaction system. This description corresponds to a simple stochastic process, namely, a continuous time Markov chain (CTMC). Consequently, in principle, we have the possibility of analysing such systems by constructing the infinitesimal generator matrix of the CTMC and computing its exponential. Computing the exponential of a matrix is not straightforward in general [17] but for matrices representing Markov chains efficient and numerically stable techniques have been developed [12,19]. Even these technique can fail however if the number of states of the Markov chain is very large or infinite, which is often the case when reaction systems are considered.

Several approximation techniques have been proposed to overcome the problem of the huge state space. The mean-field approach, based on the rela-

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

tion of the trajectories followed by the CTMC and the differential equation-based description of the system [13,14], provides a deterministic approximation of the system behaviour. The approximation provided by the mean-field approach can be seen as the approximate average behaviour of the model. The idea can be extended to higher order and joint moments of the population levels leading to more precise approximations [8].

Other techniques obtain approximations by operating directly on the state space of the model. As the state space can be infinite it is natural to bound the set of states that are considered [7]. Moreover, as the system evolves, it can be necessary to apply dynamic bounding in order to take into account at any transient time the most probable region of the state space. As the calculations can be slow even on the reduced state space, recently, faster approximate uniformisation methods have been proposed [16,23]. A further possibility to deal with the huge state space consists in aggregation. One natural option is to aggregate nearby states [22,5] while a somewhat more intricate aggregation can be obtained by applying flow equivalence [4,6].

Another approach to the analysis of large (or infinite) Markov chains representing reaction networks is simulation. Because of the huge state space and the fact that a large amount of reactions can occur in a short time interval, even simulation is not straightforward. Starting from [9], several papers have proposed approaches to increase the efficiency of simulation of reaction systems. The most used among these approaches are explicit [11] and implicit [18] tau-leaping, which uses an approximation to "leap over" many reactions in a single step, and the slow-scale stochastic simulation algorithm [3], which aims at facing stiffness of the dynamics of the model by distinguishing fast and slow reactions.

In this paper we present a novel approximation technique. We leave the state space of the model unchanged (i.e., we do not perform reductions and aggregations) but we simplify the analysis by making assumptions on the transient probabilities of the model. In particular, we assume that the transient probabilities are of product form. This assumption allows us to have a compact description of the transient probabilities and hence to analyse systems whose state space would otherwise be of prohibitive size.

The result most related to our technique is the one presented in [2] where the authors provide a necessary and sufficient condition for a network of Markovian queues to have transient product form. The condition is that the network is composed of infinite server queues. Clearly, not all reaction networks correspond to a network of infinite server queues and hence the product form assumption does not hold in general. We will show however, through numerical experiments, that the closer the reaction network is to a queueing network of infinite servers, the better the approximation will be.

The paper is organised as follows. In Section 2 the considered system of

reactions are described in brief. In Section 3 we introduce our approximation approach. Properties of the approximation are discussed in Section 4. Numerical examples are provided in Section 5. In Section 6 conclusions are drawn.

2 Model formulation

We consider a system of M reactants (called also species), $R_1, R_2, ..., R_M$, interacting through N reactions:

$$\sum_{m=1}^{M} a_{nm} R_m \xrightarrow{\lambda_n} \sum_{m=1}^{M} b_{nm} R_m, \quad 1 \leq n \leq N. \tag{1}$$

The nth reaction uses up a_{nm} units of reactant R_m and produces b_{nm} units of it. Both a_{nm} and b_{nm} are non-negative integer values and will be organised into vectors as $a_n = |a_{n1}, ..., a_{nM}|$ and $b_n = |b_{n1}, ..., b_{nM}|$. We will denote by $c_{nm} = b_{nm} - a_{nm}$ the overall effect of reaction n on reactant m and the corresponding vector will be denoted by $c_n = |c_{n1}, ..., c_{nM}|$. The speed of the nth reaction is given by λ_n. There are two classical approaches to associate a temporal behaviour with the reactions in (1).

The stochastic approach associates a continuous time Markov chain (CTMC) with the system [9]. The CTMC is discrete state, i.e., the quantity of a given reactant at any time t is given by an integer and the state of the chain is given by a vector of integers. In state $X = |X_1, ..., X_M|$ reaction n is possible if $X_m \geq a_{nm}, 1 \leq m \leq M$. We will apply the relation $\geq$ to vectors meaning that $X \geq a_n$ if and only if $X_m \geq a_{nm}, 1 \leq m \leq M$. If reaction n is possible in state X then its intensity is given by

$$\lambda_n \prod_{m=1}^{M} \binom{X_m}{a_{nm}} \tag{2}$$

and the corresponding transition takes the CTMC from state X to state $X+c_n$. We will denote by $p(X,t)$ the probability that the CTMC is in state X at time t and this quantity satisfies the following well-known Chapman-Kolmogorov ordinary differential equation (ODE)

$$\begin{aligned} \frac{dp(X,t)}{dt} = &- p(X,t) \sum_{n:X \geq a_n} \lambda_n \prod_{m=1}^{M} \binom{X_m}{a_{nm}} + \\ &\sum_{n:X-c_n \geq a_n} p(X-c_n,t) \lambda_n \prod_{m=1}^{M} \binom{X_m - c_{nm}}{a_{nm}} . \end{aligned} \tag{3}$$

The second, deterministic approach describes the reactions not as discrete transitions but by assuming that the reactions are modifying infinitesimal quantities of the involved species. Consequently, the quantities of the reactants are given by continuous values and ODEs describe the evolution of the system. By applying mass action kinetics [20], the ODEs are

$$\frac{R_m(t)}{dt} = \sum_{n=1}^{N} \lambda_n c_{nm} \prod_{i=1}^{M} \frac{R_i(t)^{a_{ni}}}{a_{ni}!} \tag{4}$$

where $R_m(t) \in \mathbb{R}_{\geq 0}$ is the quantity of reactant R_m at time t. Having set the initial values, $R_m(0), 1 \leq m \leq M$, the set of ODEs in (4) can be solved by numerical integration and results in a deterministic temporal behaviour.

The deterministic approach described above happens to be the mean-field approximation of the CTMC of the stochastic approach. Moreover, Kurtz has shown [14] that, as the initial population levels are increased and the reaction intensities are adjusted accordingly (giving rise to a series of so-called level-dependent Markov chains), the trajectory followed by the CTMC tends to the trajectory described by the ODEs of the deterministic approach.

Throughout the paper we will use the well-known Lotka-Volterra model to illustrate the approach. This model, proposed independently by Lotka [15] and Volterra [21], uses three reactions to describe the evolution of two populations in competition. The three reactions are

$$\begin{aligned} \text{growth of prey: } & R_1 \xrightarrow{\lambda_1} 2R_1, \\ \text{growth of predator: } & R_1 + R_2 \xrightarrow{\lambda_2} 2R_2, \\ \text{death of predator: } & R_2 \xrightarrow{\lambda_3} \emptyset \end{aligned}$$

and the Markov chain of the model is depicted in Figure 1. The mean-field approximation of the model is provided by the ODEs

$$\frac{R_1(t)}{dt} = \lambda_1 R_1(t) - \lambda_2 R_1(t) R_2(t), \quad \frac{R_2(t)}{dt} = \lambda_2 R_1(t) R_2(t) - \lambda_3 R_2(t)$$

which leads to oscillation along closed curves except if the system is started in equilibrium state.

3 Product form approximation

In order to derive the proposed approximation, we assume that the transient probabilities of the model are of product form. This means that, denoting by $p(x, m, t)$ the probability that at time t there are x units of reactant m, the transient probability of a state can be written as $p(X, t) = \prod_{m=1}^{M} p(X_m, m, t)$.

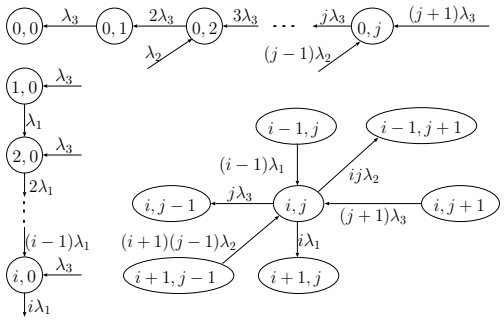

Fig. 1. Boundaries and a generic state of the Markov chain of the Lotka-Volterra model (each state is labelled by the number of preys and predators).

The quantity $p(x, m, t)$ satisfies the ODE

$$\frac{dp(x,m,t)}{dt} = \frac{d\sum_{X:X_m=x} p(X,t)}{dt} = \sum_{X:X_m=x} \frac{dp(X,t)}{dt} \tag{5}$$

which by applying (3) and the product form assumption becomes

$$\frac{dp(x,m,t)}{dt} = \sum_{X:X_m=x} \left[- \sum_{n:X\geq a_n} \lambda_n \prod_{i=1}^{M} p(X_i,i,t) \binom{X_i}{a_{ni}} + \right.$$
$$\left. \sum_{n:X-c_n\geq a_n} \lambda_n \prod_{i=1}^{M} p(X_i - c_{ni},i,t) \binom{X_i - c_{ni}}{a_{ni}} \right] . \tag{6}$$

The order of the summation in (6) can be inverted which leads to

$$\frac{dp(x,m,t)}{dt} = - \sum_{n:x\geq a_{nm}} \lambda_n \binom{x}{a_{nm}} p(x,m,t) \prod_{i=1,i\neq m}^{M} f(i,a_{ni},t) +$$
$$\sum_{n:x-c_{nm}\geq a_{nm}} \lambda_n \binom{x-c_{nm}}{a_{nm}} p(x-c_{nm},m,t) \prod_{i=1,i\neq m}^{M} f(i,a_{ni},t) \tag{7}$$

where the quantity $f(i, j, t)$ is strongly related to the jth factorial moment of the quantity of the ith reactant at time t and is defined as

$$f(i,j,t) = \sum_{k=j}^{\infty} \binom{k}{j} p(k,i,t) . \tag{8}$$

4 Properties of the approximation

The ODEs in (7) describing our product form approximation can be interpreted as M interacting, time inhomogeneous Markov chains in which a chain

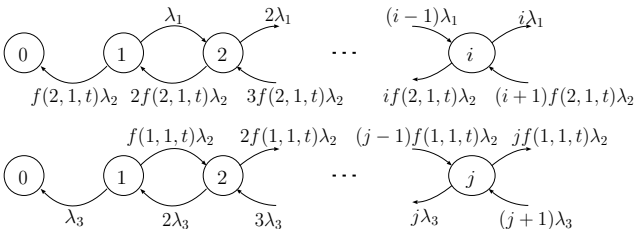

Fig. 2. Approximating time inhomogeneous Markov chains for the Lotka-Volterra model; upper part: preys, lower part: predators.

corresponding to a given reactant models the reactions in which the reactant is involved by taking into account the quantities given in (8) of the other chains. This interpretation is depicted in Figure 2 for the Lotka-Volterra model. It follows that numerical solution techniques developed for time inhomogeneous Markov chains, like the one proposed in [1], can be applied to calculate the transient probabilities.

Let us denote by $r_m, 1 \leq m \leq M$ the number of values of x for which the probability $p(x, m, t)$ is not negligible. Then the number of ODEs describing the approximation is $\sum_{m=1}^{M} r_m$. This quantity grows linearly with the number of reactants and hence the method scales well.

As mentioned in Section 2, if we consider a series of level-dependent Markov chains with increasing initial state, the transient behaviour tends to the mean-field approximation of the model [13,14]. The same holds for the approximation we proposed. Increasing initial population levels gives rise to a series of level-dependent, interacting, time inhomogeneous Markov chains and the transient behaviour for this series tends to the mean-field approximation. This means that the same relation holds between the proposed product form approximation and the mean-field approximation as between the original Markov chain model and the mean-field approximation.

5 Numerical examples

As a first example we consider an exceedingly simple model to show when the approximation fails to provide accurate results. The model is composed of the single reaction $R_1 + 2R_2 \xrightarrow{1} \emptyset$ and the starting state is $(100, 100)$. It is obvious that the model satisfies the invariant $2(100 - R_1) = 100 - R_2$ and hence the state space is composed of 51 states which are $(100, 100), (99, 98), (98, 96), \ldots, (50, 0)$. The two interactive, time inhomogeneous Markov chains representing our product form approximation is depicted in Figure 3. There are two important differences between the original model and its approximation. The approximation does not maintain the invariant of the original model. Moreover, in the original model the quantity of reactant R_1 cannot decrease below 50 while it can happen in the approximating model.

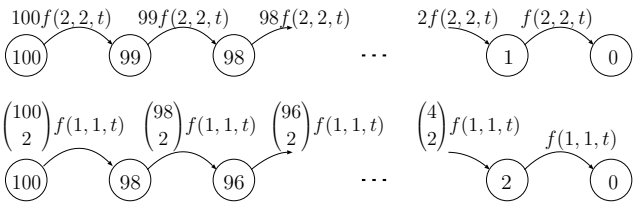

Fig. 3. Approximating time inhomogeneous Markov chains for the $R_1 + 2R_2 \xrightarrow{1} \emptyset$ model; upper part: R_1, lower part: R_2.

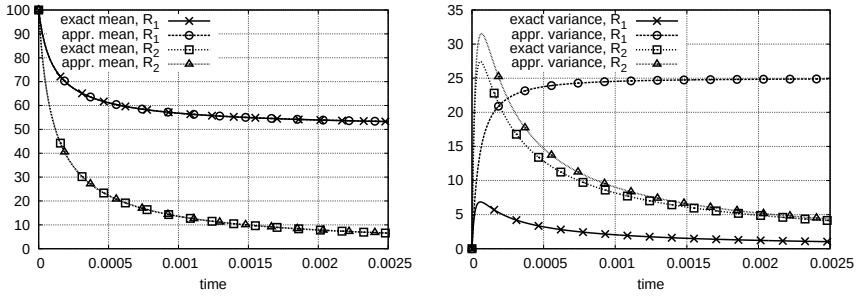

Fig. 4. Mean (left) and variance (right) for the $R_1 + 2R_2 \xrightarrow{1} \emptyset$ model.

Consequently, the approximation works with 152 differential equations and requires more computation than the original model. This is, however, not the case in general. In case of models containing more reactants and not having very restrictive invariants, the approximation requires much less calculation than the original model.

In Figure 4 we depicted the exact and the approximated mean and variance of the involved quantities as function of time. The approximation describes precisely the mean of both reactants and provides good estimate of the variance of R_2 while it fails on the variance of R_1. Figure 5 shows the distribution of the quantity of the reactants for a few different transient times. For distributions as well, the approximation is good for R_2 and it is bad for R_1. The result is not surprising as the model is far from being product form and for R_1 it introduces values which are not possible in the original model. In Figure 6 we compare the precision of the mean obtained by the product form approximation and the precision of the mean-field approach. It can be seen that the approach we proposed, even if the model is unfavourable for it, provides more precise mean values than the mean-field approximation.

As a second example we consider the preys-predators model of Lotka and Volterra. First, we start the model in state $(2000, 2000)$ and use reaction rates $\lambda_1 = 10, \lambda_2 = 0.01, \lambda_3 = 10$. The mean and the variance of the number of predators obtained by simulation and by the product form approximation are depicted in Figure 7. Even if the population levels are high, the mean deviates away soon from the stable oscillation pattern. The product form approximation predicts instead stable oscillation of the mean and provides

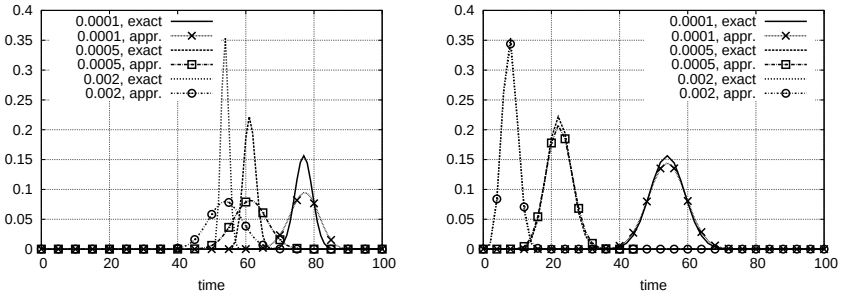

Fig. 5. The $R_1 + 2R_2 \xrightarrow{1} \emptyset$ model: distribution of R_1 (left) R_2 (right) for different transient times (for R_2 every odd value is of zero probability; in order to have a clean figure we plotted only non-zero probabilities).

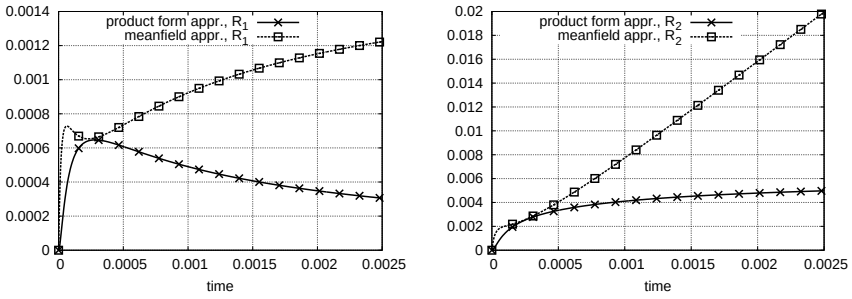

Fig. 6. The $R_1 + 2R_2 \xrightarrow{1} \emptyset$ model: relative error of the mean provided by the proposed product form approximation and of the mean-field approximation; left R_1, right: R_2.

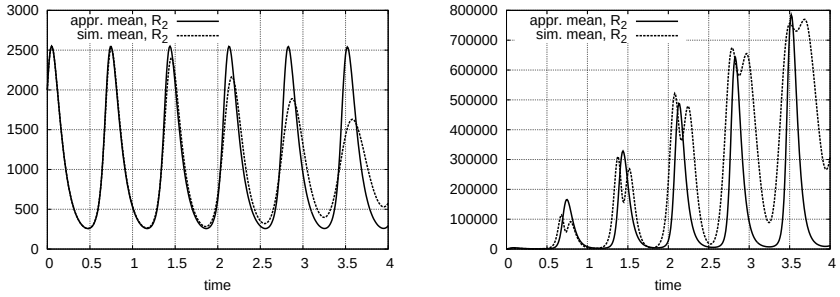

Fig. 7. The Lotka-Volterra model with $\lambda_1 = 10, \lambda_2 = 0.01, \lambda_3 = 10$ and initial state $(2000, 2000)$: mean value (left) and variance (right) of the number of predators by simulation and by product form approximation.

very similar mean to that of the mean-field approach (which is not depicted in Figure 7 because it cannot be distinguished from the mean provided by the product form approach). The variance pattern provided by the product form approximation gives instead a more precise picture of what happens in the original model.

In Figure 8 we depicted the same quantities as in Figure 7 but starting the model from state $(200, 200)$ and with rates $\lambda_1 = 10, \lambda_2 = 0.1, \lambda_3 = 10$. The mean-field approximation of the model with these parameters is the same as with the parameters used before. However, as the number of preys and

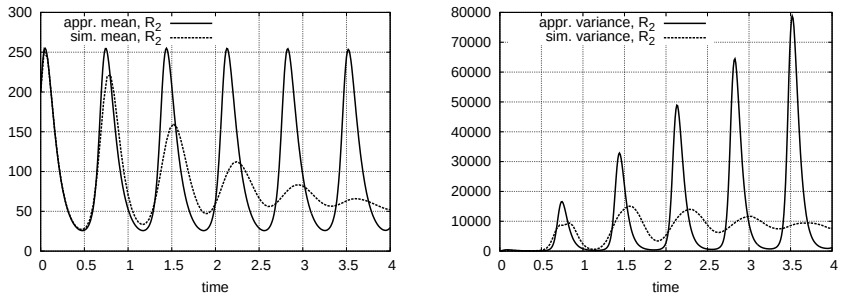

Fig. 8. The Lotka-Volterra model with $\lambda_1 = 10, \lambda_2 = 0.1, \lambda_3 = 10$ and initial state $(200, 200)$: mean value (left) and variance (right) of the number of predators by simulation and by product form approximation.

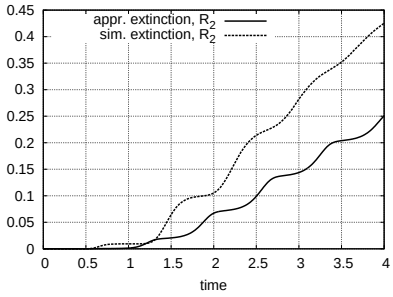

Fig. 9. The Lotka-Volterra model with $\lambda_1 = 10, \lambda_2 = 0.1, \lambda_3 = 10$ and initial state $(200, 200)$: extinction of predators by simulation and by product form approximation.

predators are lower, the model deviates from stable oscillation faster. The product form approximation is not able to capture this behaviour and provides imprecise estimate of both the mean and the variance. Figure 9 shows the probability of extinction of predators as function of time obtained by simulation and the proposed approximation. The reason that the approximation fails to give precise estimates is that the behaviour of the system depends strongly on the correlation of the population levels which is not captured by the product form probabilities.

Our third example is an extended version of the Lotka-Volterra model. We consider two types of prey and two types of predators. The set of reaction are the following:

$$\begin{aligned}
\text{growth of preys: } & R_1 \xrightarrow{10} 2R_1,\ R_2 \xrightarrow{10} 2R_2, \\
\text{growth of predators: } & R_1 + R_3 \xrightarrow{0.015} 2R_3,\ R_1 + R_4 \xrightarrow{0.03} 2R_4, \\
& R_2 + R_3 \xrightarrow{0.02} 2R_3,\ R_2 + R_4 \xrightarrow{0.025} 2R_4, \\
\text{death of predators: } & R_3 \xrightarrow{10} \emptyset,\ R_4 \xrightarrow{15} \emptyset\ .
\end{aligned}$$

The mean-field approach associates stable oscillation with model. This model, as there are more species and more reactions than in the original Lotka-Volterra model, maintains the oscillation for more time. In Figure 10 and

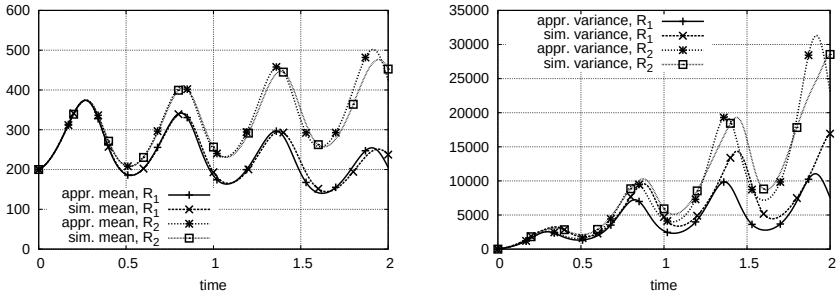

Fig. 10. The extended Lotka-Volterra model: mean (left) and variance (right) of the number of both types of preys by simulation and by product form approximation.

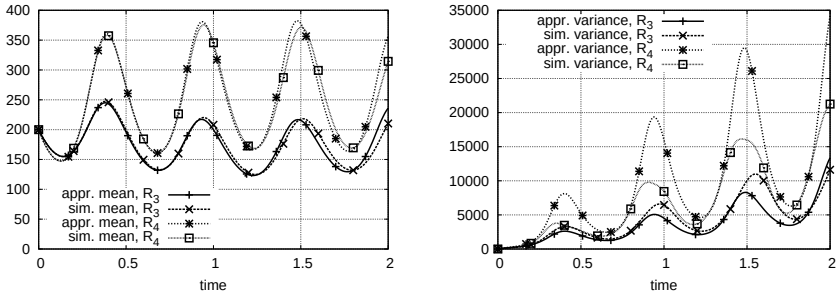

Fig. 11. The extended Lotka-Volterra model: mean (left) and variance (right) of the number of both types of predators by simulation and by product form approximation.

11 we show the mean and the variance of all the species involved in the system. The mean provided by the mean-field approach is not shown as it is indistinguishable from the mean provided by the product form approximation. For this model the product form approach gives good estimate of the mean and illustrate well the behaviour of the variance of the species. Note that, since there are four species involved, the original Markov chain is huge even if the states with negligible probability are not considered. The product form approach requires instead to solve a system of ODEs with about 6000 equations which took about 2 minutes on an ordinary laptop using the odeToJava package [1] .

As the last example we consider models which are close to networks of infinite server queues and hence their transient probabilities are approximated well in product form. We consider the reactions

$$\emptyset \xrightarrow{20} R_1, R_1 \xrightarrow{0.5} R_2, \emptyset \xrightarrow{2} R_3, R_2 \xrightarrow{1} R_3, R_3 \xrightarrow{1} R_4, R_4 \xrightarrow{1} \emptyset, R_2 + R_4 \xrightarrow{\lambda} \emptyset$$

where only the last reaction causes the system not to be a network of infinite servers. The larger the rate associated with this reaction the worse the product form approximation. The initial state is $|0, 0, 0, 0|$. In Figures 12 and 13 we show the mean and the variance of the quantity of R_2 and R_4. The mean

[1] Available at *http://www.netlib.org/ode/* and developed by M. Patterson and R. J. Spiteri.

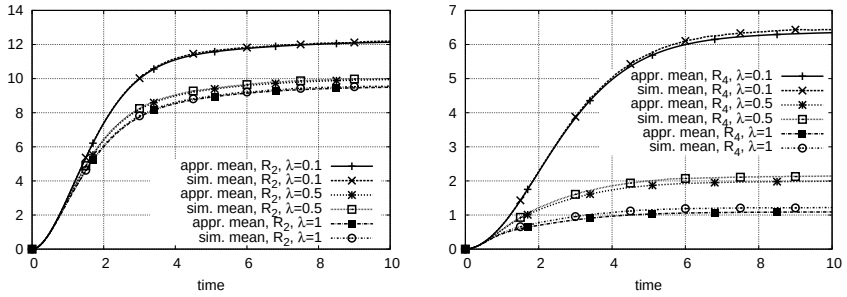

Fig. 12. Perturbed network of infinite queues: mean of R_2 (left) and R_4 (right).

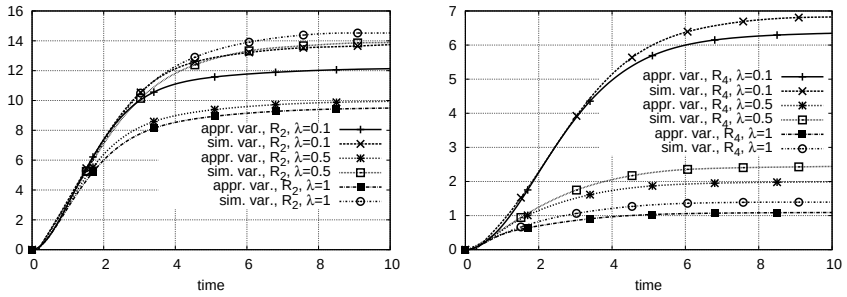

Fig. 13. Perturbed network of infinite queues: variance of R_2 (left) and R_4 (right).

is approximated well for both reactants while the variance is captured well for R_4 and is underestimated for R_2. The calculations required less than two seconds.

6 Conclusions

In this paper we derived a product form approximation for the transient probabilities of Markov chains representing reaction networks. The computational effort of the method grows only linearly with the number of reactants and hence it can be applied to reaction networks for which the exact analysis of the corresponding Markov chain is unfeasible. We tested the method on several examples and found that if the average behaviour of the system is captured well by the mean-field approach or the transient system behaviour is close to product form then the proposed approximation provides a good picture of the variance and the distribution of the quantity of the reactants.

References

[1] Arns, M., P. Buchholz and A. Panchenko, *On the numerical analysis of inhomogeneous continuous-time Markov chains*, Informs Journal on Computing **22** (2010), pp. 416–432.

[2] Boucherie, R. J. and P. Taylor, *Transient product form distributions in queueing networks*, Discrete Event Dynamic Systems: Theory and Applications **3** (1993), pp. 375–396.

[3] Cao, Y., D. T. Gillespie and L. R. Petzold, *The slow-scale stochastic simulation algorithm*, J Chem Phys **122** (2005).

[4] Chandy, K. M., U. Herzog and L. S. Woo, *Parametric analysis of queueing networks*, IBM Journal of R. & D. **19** (1975), pp. 36–42.

[5] Ciocchetta, F., A. Degasperi, J. Hillston and M. Calder, *Some investigations concerning the ctmc and the ode model derived from bio-pepa.*, Electron. Notes Theor. Comput. Sci. **229** (2009), pp. 145–163.

[6] Cordero, F., A. Horváth, D. Manini, L. Napione, M. D. Pierro, S. Pavan, A. Picco, A. Veglio, M. Sereno, F. Bussolino and G. Balbo, *Simplification of a complex signal transduction model by the application of invariants and flow equivalent server*, Submitted to Theoretical Computer Science (2011).

[7] Dayar, T., L. Mikeev and V. Wolf, *On the numerical analysis of stochastic Lotka-Volterra models*, in: *Proc. of the Workshop on Computer Aspects of Numerical Algorithms (CANA10)*, 2010, pp. 289–296.

[8] Engblom, S., *Computing the moments of high dimensional solutions of the master equation*, Appl. Math. Comput **180** (2006), pp. 498–515.

[9] Gillespie, D. T., *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81** (1977), pp. 2340–2361.

[10] Gillespie, D. T., *A rigorous derivation of the chemical master equation*, Physica A **188** (1992), pp. 404–425.

[11] Gillespie, D. T., *Approximate accelerated stochastic simulation of chemically reacting systems*, J Chem Phys **115** (2001), pp. 1716–1733.

[12] Jensen, A., *Markoff chains as an aid in the study of Markoff processes*, Skandinavisk Aktuarietidskrift **36** (1953), pp. 87–91.

[13] Kurtz, T. G., *Solutions of ordinary differential equations as limits of pure jump Markov processes*, Journal of Applied Probability **1** (1970), pp. 49–58.

[14] Kurtz, T. G., *The Relationship between Stochastic and Deterministic Models for Chemical Reactions*, J Chem Phys **57** (1972), pp. 2976–2978.

[15] Lotka, A. J., "Elements of Mathematical Biology," Williams and Wilkins Company, 1924.

[16] Mateescu, M., V. Wolf, F. Didier and T. A. Henzinger, *Fast adaptive uniformisation of the chemical master equation*, IET systems biology **4** (2010), pp. 441–452.

[17] Moler, C. and C. V. Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Review **45** (2003), pp. pp. 3–49.

[18] Rathinam, M., L. R. Petzold, Y. Cao and D. T. Gillespie, *Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method*, J Chem Phys **119** (2003), pp. 12784–12794.

[19] Stewart, W. J., "Introduction to the Numerical Solution of Markov Chains." Princeton University Press, 1995.

[20] Voit, E. O., "Computational Analysis of Biochemical Systems," Cambridge University Press, 2000.

[21] Volterra, V., *Fluctuations in the abundance of a species considered mathematically*, Nature (1924).

[22] Zhang, J., L. T. Watson and Y. Cao, *Adaptive aggregation method for the chemical master equation*, Int. J. of Computational Biology and Drug Design **2** (2009), pp. 134–148.

[23] Zhang, J., L. T. Watson and Y. Cao, *A modified uniformization method for the solution of the chemical master equation*, Computers & Mathematics with Applications **59** (2010), pp. 573 – 584.