Quantitative Methods in Systems Biology
Part III: Simulation and Model-Checking

Stephen Gilmore
LFCS, University of Edinburgh

SFM:08-Bio Summerschool
Bertinoro, Italy

2nd June 2008

Outline

Acknowledgements

F. Ciocchetta, S. Gilmore, M.L. Guerriero, J. Hillston.

Stochastic Simulation and Probabilistic Model-Checking for the Analysis of Biochemical Systems.

Submitted for publication,

2008.

Probabilistic model-checking

- Probabilistic model-checking can provide valuable insight into the behaviour of biochemical systems, answering quantitative queries which cannot be addressed by stochastic simulation.

Probabilistic model-checking

- Probabilistic model-checking can provide valuable insight into the behaviour of biochemical systems, answering quantitative queries which cannot be addressed by stochastic simulation.
- However, it is a computationally intensive technique which can become infeasible if the system under consideration is too large.

Probabilistic model-checking

- Probabilistic model-checking can provide valuable insight into the behaviour of biochemical systems, answering quantitative queries which cannot be addressed by stochastic simulation.
- However, it is a computationally intensive technique which can become infeasible if the system under consideration is too large.
- Moreover, the finite nature of the state representation used means that *a priori* bounds must be set for the numbers of molecules of each species to be observed in the system.

Probabilistic model-checking

- Here we present an approach which addresses these problems by using stochastic simulation and the PRISM probabilistic model checker in tandem.

Probabilistic model-checking

- Here we present an approach which addresses these problems by using stochastic simulation and the PRISM probabilistic model checker in tandem.
- The stochastic simulation identifies reasonable bounds for molecular populations in the context of the considered experiment.

Probabilistic model-checking

- Here we present an approach which addresses these problems by using stochastic simulation and the PRISM probabilistic model checker in tandem.
- The stochastic simulation identifies reasonable bounds for molecular populations in the context of the considered experiment.
- These bounds are used to parameterise the PRISM model and limit its state space.

Probabilistic model-checking

- Here we present an approach which addresses these problems by using stochastic simulation and the PRISM probabilistic model checker in tandem.
- The stochastic simulation identifies reasonable bounds for molecular populations in the context of the considered experiment.
- These bounds are used to parameterise the PRISM model and limit its state space.
- Crucially, the technique quantifies an estimation for the truncation error incurred.

Outline

Motivation

- Analysing models of biological processes via probabilistic model-checking has considerable appeal.

Motivation

- Analysing models of biological processes via probabilistic model-checking has considerable appeal.
- As with stochastic simulation the answers which are returned from model-checking give a thorough stochastic treatment to the small-scale phenomena which are of greatest interest to computational biologists today.

Motivation

- Analysing models of biological processes via probabilistic model-checking has considerable appeal.
- As with stochastic simulation the answers which are returned from model-checking give a thorough stochastic treatment to the small-scale phenomena which are of greatest interest to computational biologists today.
- However, in contrast to a simulation run which generates just one of many possible trajectories, the analysis results computed by probabilistic model-checking give a definitive answer.

Motivation

- Analysing models of biological processes via probabilistic model-checking has considerable appeal.
- As with stochastic simulation the answers which are returned from model-checking give a thorough stochastic treatment to the small-scale phenomena which are of greatest interest to computational biologists today.
- However, in contrast to a simulation run which generates just one of many possible trajectories, the analysis results computed by probabilistic model-checking give a definitive answer.
- That is, it is not necessary to re-run the analysis repeatedly and compute ensemble averages of the results.

Motivation

- Further, by building a reward structure over the model it is possible to express complex analysis questions and evaluate these through model-checking.

Motivation

- Further, by building a reward structure over the model it is possible to express complex analysis questions and evaluate these through model-checking.
- This form of analysis has the power to expose of the system under study significant temporal behaviour which could not be appreciated from simple inspection of the species time-series generated by simulation runs.

Problem: state-space explosion

- Set against this, probabilistic model-checking of stochastic models of reacting biochemical species described at a molecular level of detail faces the well-known problem of *state-space explosion* where, as the complexity of the system under study increases, there is an exponential growth in the state-space of the underlying model.

Problem: state-space explosion

- Set against this, probabilistic model-checking of stochastic models of reacting biochemical species described at a molecular level of detail faces the well-known problem of *state-space explosion* where, as the complexity of the system under study increases, there is an exponential growth in the state-space of the underlying model.
- The use of an exact discrete-state representation of the state-space of the model restricts the use of probabilistic model-checking to the analysis of problems where all of the species are available in low copy numbers.

Bounds on species

- Even in the case where all of the chemical species involved are present only in low copy numbers it is still necessary to place a bound on the maximum molecule count which each species will attain.

Bounds on species

- Even in the case where all of the chemical species involved are present only in low copy numbers it is still necessary to place a bound on the maximum molecule count which each species will attain.
- For models involving biochemical processes such as synthesis no such bounds can be established.

Bio-PEPA model of Michaelis-Menten

$$\begin{aligned} f_{r_1} &= k_1 \times E \times S \\ f_{r_{-1}} &= k_{-1} \times E{:}S \\ f_{r_2} &= k_2 \times E{:}S \end{aligned}$$

Bio-PEPA model of Michaelis-Menten

$$\begin{aligned} f_{r_1} &= k_1 \times E \times S \\ f_{r_{-1}} &= k_{-1} \times E{:}S \\ f_{r_2} &= k_2 \times E{:}S \end{aligned}$$

$$\begin{aligned} E &\stackrel{def}{=} (r_1, 1)\downarrow E + (r_{-1}, 1)\uparrow E + (r_2, 1)\uparrow E; \\ S &\stackrel{def}{=} (r_1, 1)\downarrow S + (r_{-1}, 1)\uparrow S; \\ E{:}S &\stackrel{def}{=} (r_1, 1)\uparrow E{:}S + (r_{-1}, 1)\downarrow E{:}S + (r_2, 1)\downarrow E{:}S; \\ P &\stackrel{def}{=} (r_2, 1)\uparrow P; \end{aligned}$$

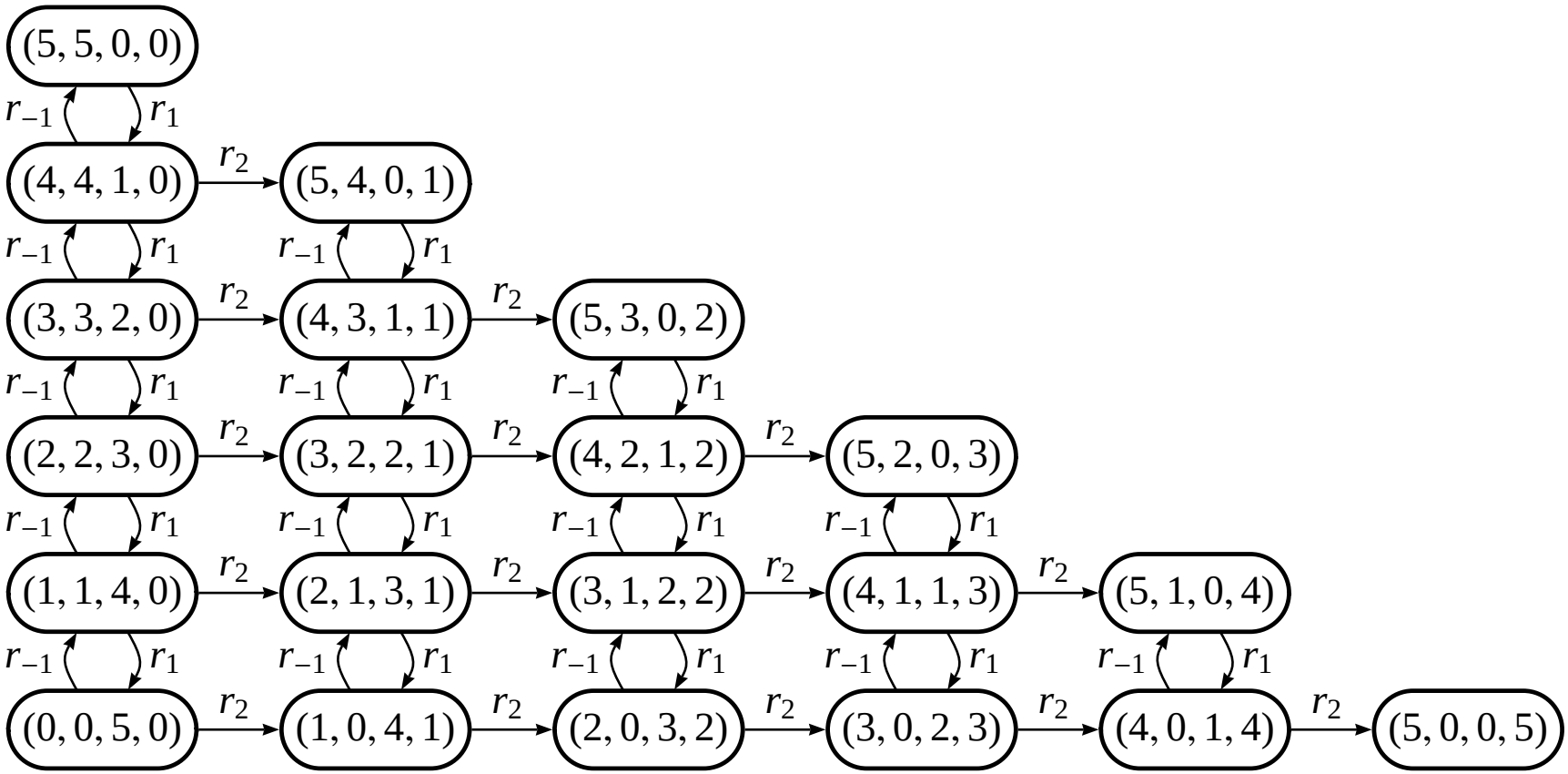
Bio-PEPA model of Michaelis-Menten

$$\begin{aligned} f_{r_1} &= k_1 \times E \times S \\ f_{r_{-1}} &= k_{-1} \times E{:}S \\ f_{r_2} &= k_2 \times E{:}S \end{aligned}$$

$$\begin{aligned} E &\stackrel{def}{=} (r_1, 1)\downarrow E + (r_{-1}, 1)\uparrow E + (r_2, 1)\uparrow E; \\ S &\stackrel{def}{=} (r_1, 1)\downarrow S + (r_{-1}, 1)\uparrow S; \\ E{:}S &\stackrel{def}{=} (r_1, 1)\uparrow E{:}S + (r_{-1}, 1)\downarrow E{:}S + (r_2, 1)\downarrow E{:}S; \\ P &\stackrel{def}{=} (r_2, 1)\uparrow P; \end{aligned}$$

$$(E(l_{E,0}) \underset{\{r_1, r_{-1}, r_2\}}{\bowtie} (S(l_{S,0}) \underset{\{r_1, r_{-1}\}}{\bowtie} (E{:}S(l_{ES,0}) \underset{\{r_2\}}{\bowtie} P(l_{P,0}))))$$

Discrete state-space of Michaelis-Menten example



Adding synthesis to the model

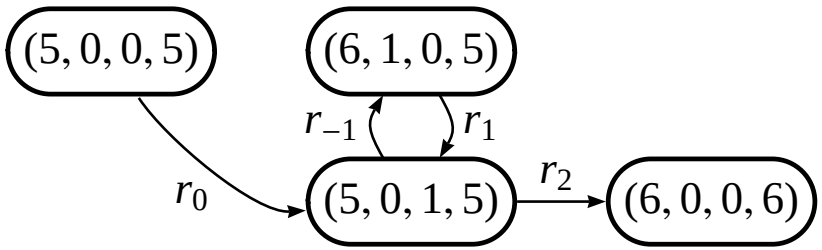
If we consider an extension of the model with an additional reaction r_0 which synthesises the compound $E{:}S$ as shown below

$$E{:}S \stackrel{def}{=} (r_0, 1)\uparrow E{:}S + (r_1, 1)\uparrow E{:}S + (r_{-1}, 1)\downarrow E{:}S + (r_2, 1)\downarrow E{:}S;$$

with the synthesis occurring at a constant rate $r_0 = k_0$ then this additional reaction channel changes the analysis of the model dramatically.

Adding synthesis to the model

The state which was previously a deadlock state now admits an r_0 reaction which leads it to a previously unreachable state, $(5, 0, 1, 5)$. The reactions r_{-1}, r_1 and r_2 can occur in states reachable from that.



Adding synthesis to the model

- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.

Adding synthesis to the model

- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.
- The effect of introducing this single synthesis reaction is that we now cannot find any upper bound N such that the molecular species counts are guaranteed to lie in the bounded integer range 0 to N.

Adding synthesis to the model

- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.
- The effect of introducing this single synthesis reaction is that we now cannot find any upper bound N such that the molecular species counts are guaranteed to lie in the bounded integer range 0 to N.
- If we are unable to bound the reachable state-space then we cannot analyse our model by probabilistic model-checking.

Observations about reachability

1. The generation of the derivation graph of the underlying state-space does not take into account the numerical values assigned to the rate constants, and the propensity functions which depend on those. This means that the derivation graph may include many states which the system is almost sure not to reach within a particular time bound.

Observations about reachability

2. Most chemical systems involve several widely varying time scales, so such systems are nearly always stiff. A consequence of this is that the first passage time to many states is likely to be long and truncation of the state-space using a time-bounded reachability metric is likely to be productive.

Observations about reachability

3. Many of the logical formulae which we wish to check involve reaching within a fixed time bound model states which satisfy a given predicate.

Observations about reachability

4. Stochastic simulation methods such as Gillespie's Direct Method generate exact stochastic simulations of trajectories from the initial state to states reachable within a given time bound.

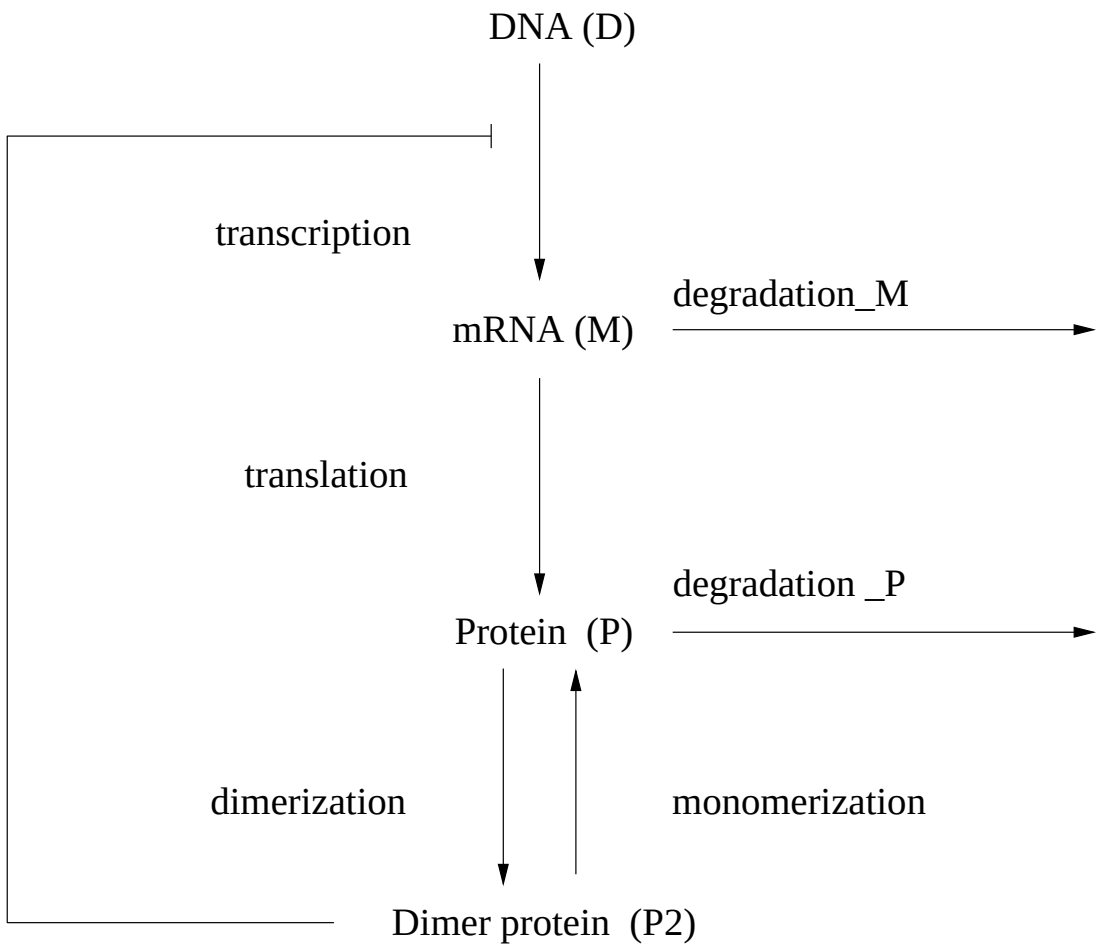
Outline

Examples: Two Genetic Networks

In order to illustrate our approach we consider two models. These represent, under different assumptions, a general genetic network with a negative feedback. An example of this kind of network is the control circuit for the λ repressor protein CI of λ-phage in *E.Coli*.

We have four biochemical entities that interact with each other through six reactions. The biochemical entities are the DNA (D), the mRNA (M), a protein in monomeric form (P) and a protein in dimeric form ($P2$).

A schema of the general network



The network is unbounded

The network is structurally unbounded, since both transcription and translation lead to the creation of new molecules.

The network is unbounded

The network is structurally unbounded, since both transcription and translation lead to the creation of new molecules.

However, the two degradation reactions and the transcription inhibition by means of the dimeric protein have a regulatory effect on the protein synthesis and therefore, under some conditions, all the species reach a finite average value.

The Network with Protein Degradation ($\mathcal{M}_1$)

We perform 1000 independent stochastic simulation runs using Gillespie's Direct Method. The number of runs is large enough to take into account the variability of the system, but still making the total simulation time reasonable. We used $T = 20000\ s$ as a simulation stop time: by that time the system has reached a stable state.

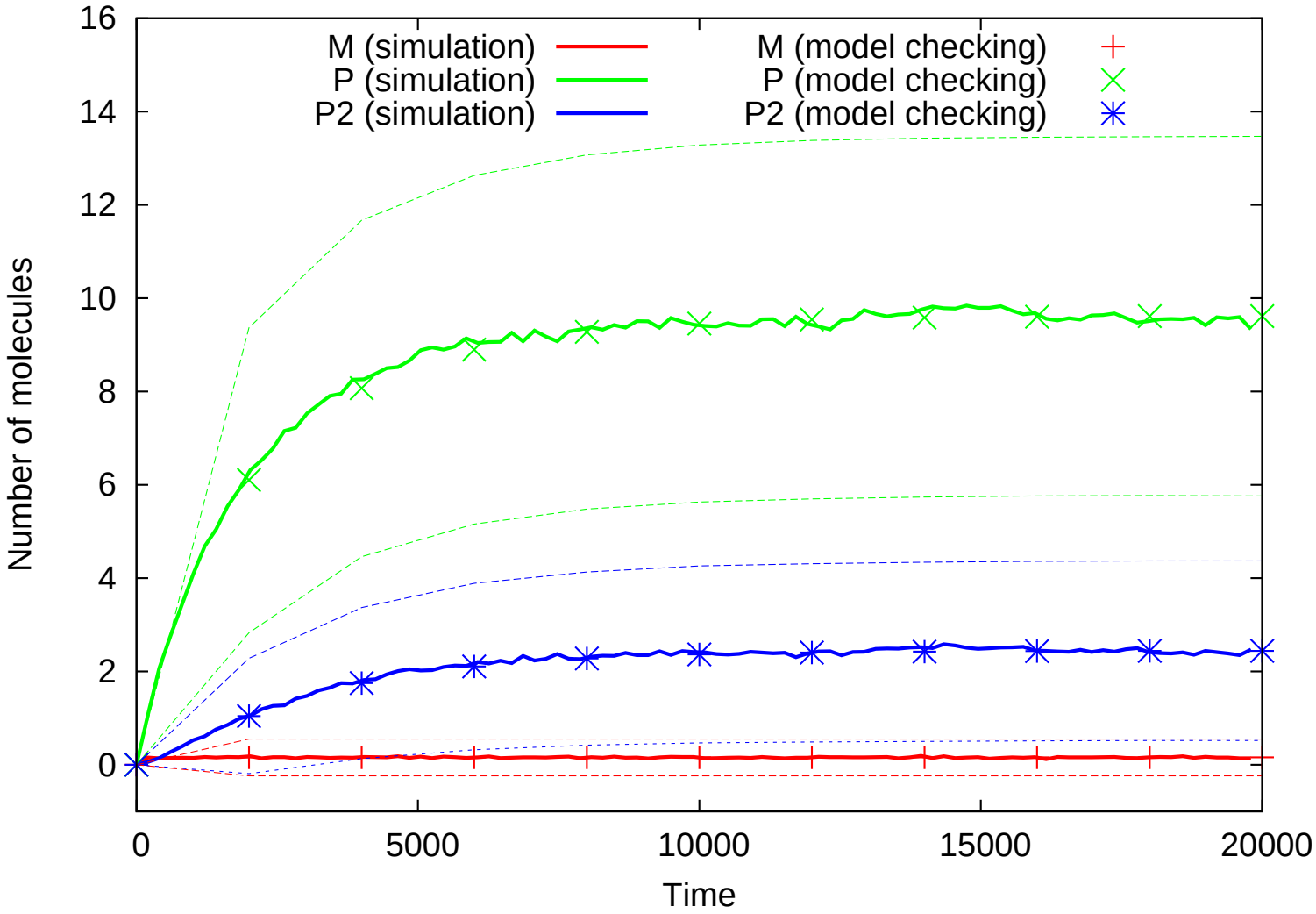
The Network with Protein Degradation ($\mathcal{M}_1$)

We can estimate the upper bounds for the amounts of each species as the maximum values obtained in any run at any time instant,

$$Max_M = 5; \qquad Max_P = 33; \qquad Max_{P2} = 18$$

and we can use these values in the PRISM model.

Simulation averages and model-checking for $\mathcal{M}_1$



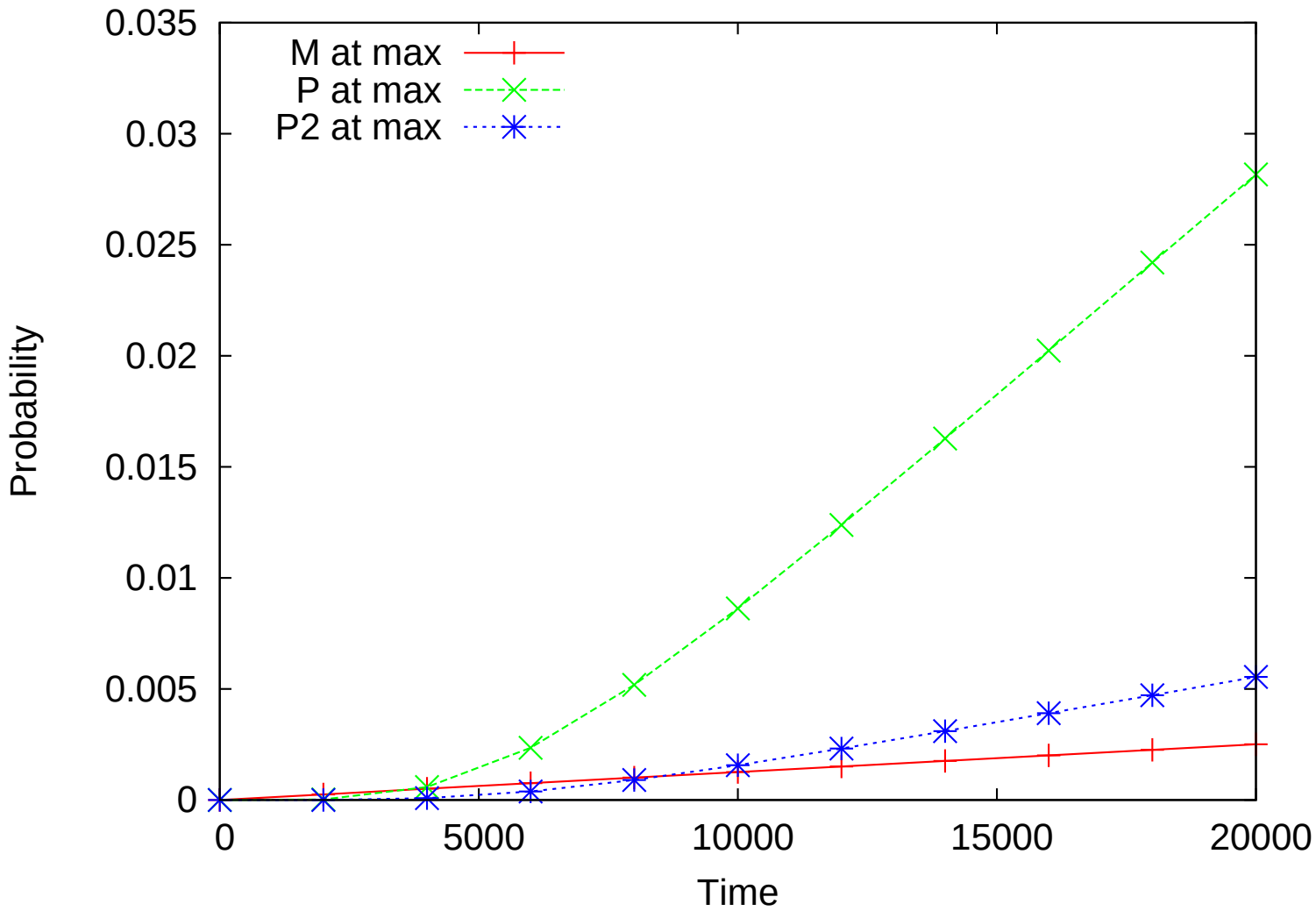
Estimating the error introduced by truncation

As another form of validation of the derived bounds, we have calculated the probabilities of reaching them at different time instants:

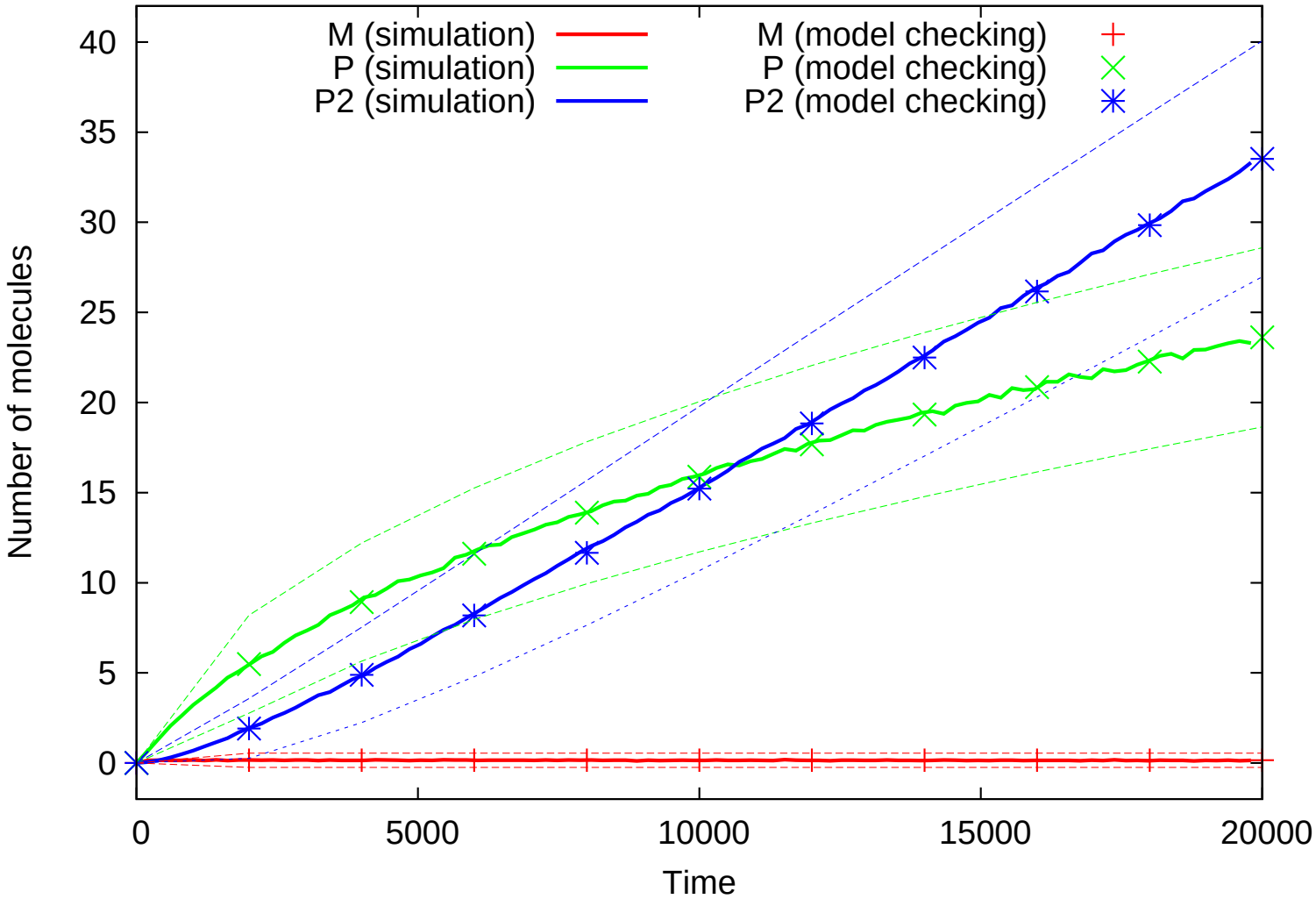
- $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ M = 5]$,
- $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ P = 33]$, and
- $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ P2 = 18]$.

The results provide a means of estimating the error which might have been introduced by bounding the system.

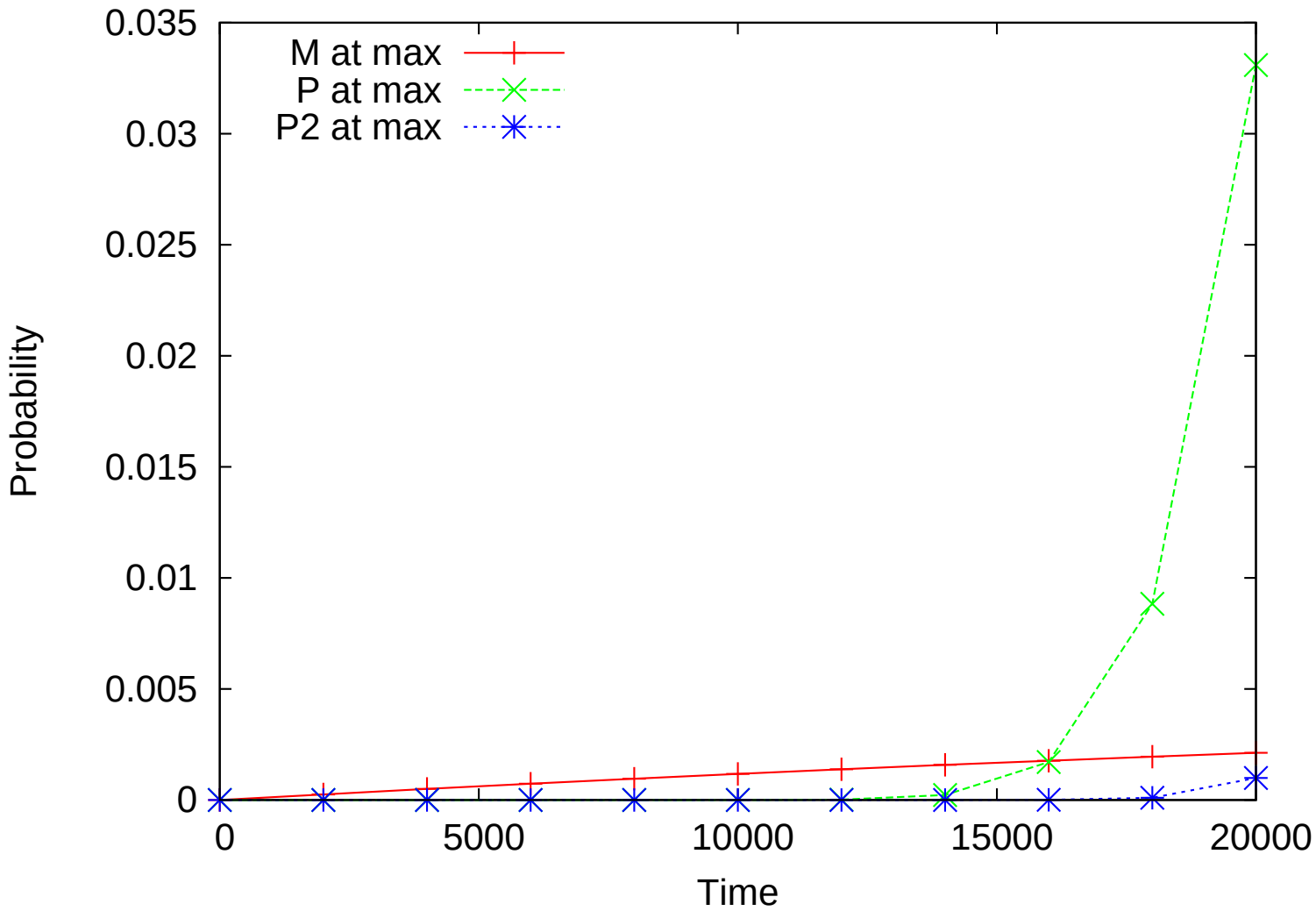
Estimating the error introduced by truncation



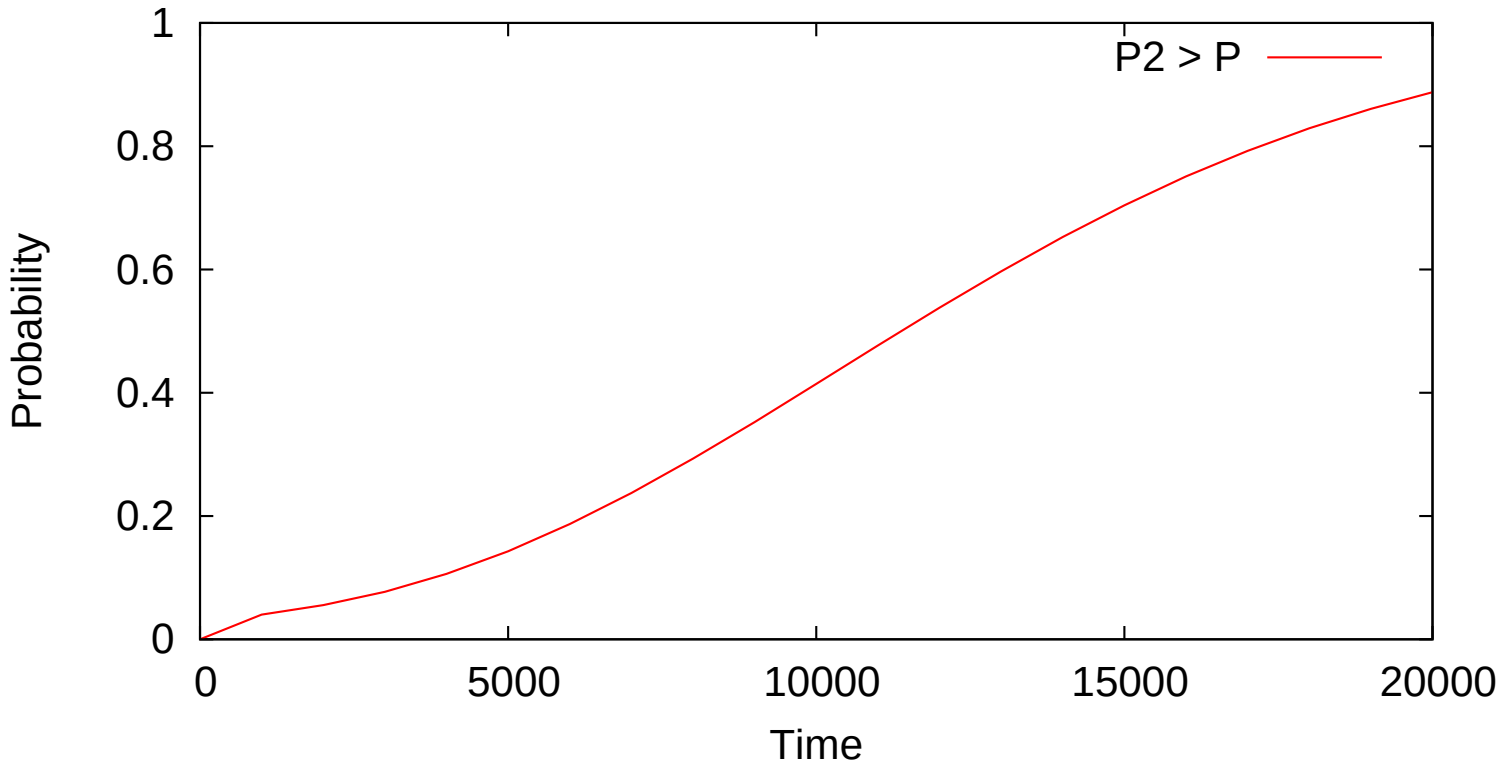
The Network Without Protein Degradation ($\mathcal{M}_2$)



Estimating the error introduced by truncation



Determining the probability that $P2 > P$



Outline

Summary

- We consider a Bio-PEPA system representing a biochemical network, and we automatically derive from it a model specification to be used for stochastic simulation and one to be used for model-checking (using PRISM).

Summary

- We consider a Bio-PEPA system representing a biochemical network, and we automatically derive from it a model specification to be used for stochastic simulation and one to be used for model-checking (using PRISM).
- We set the simulation time T and the number of simulation runs.

Summary

- We consider a Bio-PEPA system representing a biochemical network, and we automatically derive from it a model specification to be used for stochastic simulation and one to be used for model-checking (using PRISM).
- We set the simulation time T and the number of simulation runs.
- We pick as bound for a species the largest number of molecules which that species has obtained in any simulation run within time T.

Summary

- We update the PRISM model derived from the Bio-PEPA model with the estimated bounds, and we validate this model by comparing the expected values calculated by PRISM with the average values obtained by simulation.

Summary

- We update the PRISM model derived from the Bio-PEPA model with the estimated bounds, and we validate this model by comparing the expected values calculated by PRISM with the average values obtained by simulation.
- We use PRISM to analyse the model by verifying specific logical properties.