Diagnostics for assessing the accuracy of approximate stochastic simulators

Colin S. Gillespie

School of Mathematics & Statistics, Newcastle University

Abstract Solving the chemical master equation exactly is typically not possible. Instead, we must rely on simulation based methods. Unfortunately, exact realisations requires simulating every reaction that occurs. Approximate simulation techniques therefore become important. We describe a general framework to assess the linear noise and two moment approximations. By constructing an efficient space filling design over the parameter region of interest, we present a number of useful diagnostic tools. In particular, we leverage the normality assumption of the linear noise and moment closure approximations.

1 Introduction

A standard approach to modelling a set of chemical reactions is to form the chemical master equation (CME). While it may not be possible to solve the CME, it is usually straightforward to obtain exact realisations using forward simulations. Unfortunately, drawing exact realisations can be computationally intensive.

One strategy for approximating a model is to utilise the moment equations[1]. These equations can be extracted by multiplying the chemical master equation by the multivariate moment generating functions. This gives a coupled set of ordinary differential equations (ODEs). To close the system, normality is usually assumed. The linear noise approximation (LNA) is similar to the 2MA approximation, although usually derived in a different way[2].

2 Strategy

Given a set of parameter ranges that we are interested in, it would be impossible to numerically assess the approximate simulator at all values. One strategy to explore the parameter space is to use Latin hypercube sampling (LHS). This is a significant improvement over simple random sampling when exploring large spaces[3]. Our general strategy is at each of the \( n_d \) points on the Latin hypercube design, we will assess the normality assumption of the LNA.
Figure 1. Model diagnostics for the Schlögl system. (a) Latin hypercube design (on the log10 scale), with \( n_d = 1000 \). (b) 95\% credible regions and associated LNA values. (c) Standardised prediction errors, with 95\% and 99.9\% regions indicated by grey lines. A lowess line smoother is also shown in blue. (d) Fifty stochastic simulations with the parameter values associated with the largest prediction error.

3 Schlögl system

The Schlögl model is a well known model that exhibits non-linear and bimodal characteristics at particular parameter combinations. The system contains three chemical species, four reactions and assumes mass action kinetics. In this example, we concentrate on species \( X_1 \). We fix \( \{c_1, c_2\} = \{3 \times 10^{-7}, 10^{-4}\} \) and \( \{X_1(0), X_2(0), X_3(0)\} = \{250, 100000, 200000\} \). At each of the \( n_d = 1000 \) points from a two dimensional Latin hypercube on the log10 space for \( \{c_3, c_4\} \), we simulate using the direct method and the LNA to \( t = 5 \). The design space is given in figure 1a.

Standardised prediction errors are obtained by subtracting the LNA mean and dividing by the standard deviation. If the normality assumptions are correct, the errors should have a standard normal distribution. Figure 1b shows the plotted against \( c_4 \) with a locally smoothed mean value. However, there are a few large errors, in particular, \( e_{55}^s \approx -8.5 \).

Figure 1c shows the 95\% credible intervals plotted against the parameter \( c_3 \). Overall, the proportion of exact simulations falling outside the prediction interval is approximately equal to \( \alpha = 0.95 \).

Fifty stochastic simulations with parameter values equal to those used in \( e_{55}^s \) are shown in figure 1d. The LNA mean solution is also shown in red. It is clear that at this particular choice of parameter values, the Schlögl system has a bimodal distribution and the LNA is inappropriate in this region of parameter space.

References