SPSens: a software package for stochastic parameter sensitivity analysis of biochemical reaction networks

Patrick W. Sheppard[1], Muruhan Rathinam[2] and Mustafa Khammash[1,*]

[1]Department of Biosystems Science and Engineering, ETH–Zurich, Basel 4058, Switzerland and [2]Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Associate Editor: Martin Bishop

ABSTRACT

Summary: SPSens is a software package for the efficient computation of stochastic parameter sensitivities of biochemical reaction networks. Parameter sensitivity analysis is a valuable tool that can be used to study robustness properties, for drug targeting, and many other purposes. However its application to stochastic models has been limited when Monte Carlo methods are required due to extremely high computational costs. SPSens provides efficient, state of the art sensitivity analysis algorithms in a single software package so that sensitivity analysis can be easily performed on stochastic models of biochemical reaction networks. SPSens implements the algorithms in C and estimates sensitivities with respect to both infinitesimal and finite perturbations to system parameters, in many cases reducing variance by orders of magnitude compared to basic methods. Included among the features of SPSens are serial and parallel command line versions, an interface with Matlab, and several example problems.

Availability: SPSens is distributed freely under GPL version 3 and can be downloaded from http://sourceforge.net/projects/spsens/. The software can be run on Linux, Mac OS X and Windows platforms.

Contact: mustafa.khammash@bsse.ethz.ch

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Stochastic models are used now more than ever to study, to analyse, and to predict the complex dynamics occurring in biology, chemistry and ecology. When the discrete and random nature of the system is not accurately captured by deterministic or stochastic differential equation models, stochastic reaction networks are often modeled in the framework of stochastic chemical kinetics (Gillespie, 1976, 1977). A stochastic chemical kinetic model cannot be solved exactly except in the simplest cases and precludes the use of simulation and analysis methods commonly available for deterministic or stochastic differential equation models. One such analysis tool that is of particular importance is parameter sensitivity analysis, which quantifies how system output variables are affected by perturbations in model parameters. The stochastic model depends upon a set of reaction rate parameters, and it is of great interest to assess the dependence of system behaviour on these models *via* parameter sensitivity analysis. Such knowledge

has wide applications, but the computation of the stochastic sensitivity coefficients is burdensome. In particular Monte Carlo (MC) methods are used, but are very often computationally demanding because they require large numbers of simulations to be performed in order to generate accurate statistics.

Although software packages have been developed for sensitivity analysis using the linear noise approximation and Fisher Information Matrix (Komorowski *et al.*, 2012), such an approach is not appropriate for systems with particularly large fluctuations in molecular populations, such as those exhibiting bistability. Several MC methods for efficiently estimating stochastic parameter sensitivities of biochemical reaction networks have been reported recently (Anderson, 2012; Plyasunov and Arkin, 2007; Rathinam *et al.*, 2010; Sheppard *et al.*, 2012). Existing software packages have been developed to increase efficiency of stochastic simulations; however, such packages do not include methods for sensitivity analysis. To perform sensitivity analysis one must either write custom code or make significant modifications to existing code. SPSens addresses this problem by including efficient, state of the art parameter sensitivity analysis methods in a single software package. The capabilities, implementation and features are described in what follows.

2 METHODS

2.1 Stochastic sensitivity analysis

In stochastic chemical kinetics, a model is defined by specifying a list of species whose populations at time $t > 0$ are described by the discrete state vector, $\mathbf{X}(t)$. These species interact with one another through a number of reaction channels, each of which is defined by a stoichiometry vector and a propensity function. The propensity functions depend on a set of reaction rate parameters, θ. The objective of sensitivity analysis is to quantify how an output function of the form $f(\mathbf{X}(t;\theta))$ changes with perturbations to the components of θ. The sensitivity coefficients estimated by SPSens are analogous to first order sensitivity coefficients commonly computed for deterministic systems. Owing to the stochastic nature of the system, sensitivities of the *expected value* of the output—$E[f(\mathbf{X}(t;\theta))]$—with respect to perturbations of a particular system parameter are considered. These are defined as,

$$z_j = \frac{\partial}{\partial \theta_j} E[\, f(\mathbf{X}(T; \theta)], \tag{1}$$

where $\theta_j \in \theta$ is the jth system parameter and $T > 0$ is a fixed time of interest. Although (1) can be derived directly from the Chemical Master Equation (CME), solutions to these equations are intractable except for the simplest examples. However, one can use MC methods and stochastic simulation algorithms (SSA) to estimate z_j.

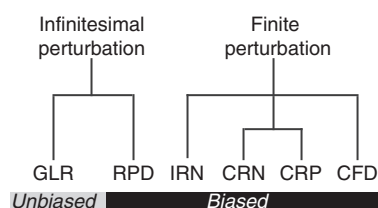**To whom correspondence should be addressed.*

Fig. 1. Monte Carlo stochastic parameter sensitivity algorithms included in SPSens

For MC methods, a *sensitivity estimator*, denoted $\hat{Z}_j$, is constructed and computed *via* stochastic simulation. A large number of samples (N) of the estimator are simulated, and the sample average, $\hat{z}_j := \frac{1}{N}\sum_{i=1}^{N}\hat{Z}_j^{(i)}$, is then taken as the sensitivity estimate. Its accuracy can be assessed by taking into account its bias (when present) and by constructing confidence intervals of the point sensitivity estimate in terms of the sample variance. Estimators with no or little bias and low variance are thus desired for the most accurate estimates. Several different sensitivity estimators have been proposed, each differing in its bias and variance properties.

Finite difference (FD) estimators are among the most straightforward and generally applicable estimators. FD estimates are computed by first simulating the system at a nominal parameter set and then at a parameter set perturbed by a small but finite amount and then taking the FD quotient. The resulting estimate given by the sample average of many FD quotient samples has a bias that depends on the particular discretization scheme implemented. The variance of the estimates depends on *how* the samples are generated. A direct approach is to generate samples using independent and identically distributed random numbers (IRN), but this estimate often has high variance and requires many samples. This variance can be reduced significantly by instead generating samples that are correlated in an appropriate manner. Using common random numbers (CRN) can reduce variance significantly to increase efficiency. Although this approach combined with the direct method SSA (Gillespie, 1977) can be easily applied to reduce variance considerably, further gains in efficiency can be achieved by using the common reaction path (CRP) method (Rathinam *et al.*, 2010). CRP implements CRN with an SSA based on the random time change representation of Markov processes to reduce variance significantly. Another approach is the coupled finite difference (CFD) method (Anderson, 2012), which simultaneously simulates coupled processes for each parameter set to yield a FD estimate with highly desirable variance properties in many situations, especially when considering larger time intervals.

Other approaches directly consider infinitesimal perturbations to the parameters to obtain the estimate. One such method, the Girsanov likelihood ratio (GLR), was first introduced in the context of stochastic chemical kinetics in (Plyasunov and Arkin, 2007). This method derives numerically tractable weighting coefficients using the Girsanov measure transformation that can be computed during simulation to give an unbiased sensitivity estimate. GLR, although unbiased, often has high variance and can be inefficient in some situations. The Regularized Pathwise Derivative (RPD) method is an alternative approach that considers a close regularization of the original problem and computes the derivatives for each sample path to estimate the sensitivities (Sheppard *et al.*, 2012). RPD estimates are often significantly more efficient than the other methods, with variance reduction of over 10 orders of magnitude demonstrated for certain problems.

A brief numerical comparison of these methods using example stochastic biochemical networks is provided in the Supplementary Material.

2.2 Implementation

SPSens implements the sensitivity algorithms summarized in Figure 1 in C and requires no external libraries outside the standard distributions. To run the program, the biochemical network of interest is first defined in a C source file by specifying the species, reactions and parameters, and additionally the partial derivatives of the propensity functions with respect to system parameters if GLR or RPD are desired. Several examples are provided with the SPSens distribution and can be easily modified for user-defined networks.

Once the problem is defined, the code is compiled to build an executable program that can be run from the command line or using a scripting language. The sensitivity analysis algorithm can be specified within the problem file or at run-time by setting the appropriate command line flag. The sampling times, parameters and initial conditions can also be specified via command-line, permitting sensitivity analysis to be performed for a variety of conditions without the need to recompile any code.

The methods using common random numbers (CRP, CRN) require special handling of the pseudo random numbers driving the SSA. Random number generation is therefore handled using the RngStream package (L'Ecuyer *et al.*, 2002). The package supports parallel random number generation streams of long period, which are quite useful for ensuring proper synchronization of random numbers and independence of parallel streams.

2.3 Other features

In addition to sensitivity analysis programs, SPSens includes exact SSA that can be used to generate individual sample paths or ensembles of trajectories using the identical model definition. The command line sensitivity analysis and simulation programs are also implemented using the Message Passing Interface (MPI), permitting the programs to be run in parallel on a grid or on multi-core personal computers.

SPSens includes the option to perform sensitivity analysis or simulation using *antithetic sampling* (AS). AS is a widely used variance reduction technique that in many cases induces negative correlation in samples to produce estimates with tighter confidence intervals. AS is easily incorporated with all algorithms in SPSens by setting a single flag from the command line. It should be noted, however, that variance reduction with AS is typically modest and that variance reduction is not guaranteed in general, so AS should be used keeping this in mind.

Finally, source code allowing SPSens C-code to be compiled and run from MATLAB using the MEX interface are included with SPSens. An example m-file script is included to demonstrate how to build and call the programs from within MATLAB.

Much more detailed information can be found in the User Guide provided with the SPSens distribution.

3 CONCLUSION

SPSens is a freely available and easy to use software package that implements efficient methods for parameter sensitivity analysis of discrete stochastic biochemical reaction networks. SPSens allows users to quickly and easily switch between algorithms to find the most efficient method for their particular problem.

In general it is impossible to say which algorithm is best for a given problem, as the performance of each method depends on a number of factors. However there are some general guidelines that can be followed in practice. GLR is the only unbiased method, but it is often very inefficient and requires significant computational effort to produce accurate results. The FD methods with IRN and CRN are significantly less efficient than the CFD and CRP methods in most situations and are provided mainly for demonstration purposes. The RPD method is extremely efficient for a wide variety of problems when applicable, but requires judicious choice of the regularization parameter for efficient performance. Which of the three highly efficient methods (RPD, CFD and CRP) is most efficient will depend on the

problem, however. Example results computed with SPSens and a more detailed discussion of these matters is provided in the Supplementary Material.

Conflict of Interest: none declared.

REFERENCES

Anderson,D. (2012) An efficient finite difference method for parameter sensitivities of continuous time markov chains. *SIAM J. Numer. Anal.*, **50**, 2237–2258.

Gillespie,D. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, **22**, 403–434.

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

Komorowski,M. *et al.* (2012) Stochsens–MATLAB package for sensitivity analysis of stochastic chemical systems. *Bioinformatics*, **28**, 731–733.

L'Ecuyer,P. *et al.* (2002) An object-oriented random-number package with many long streams and substreams. *Oper. Res.*, **50**, 1073–1075.

Plyasunov,S. and Arkin,A. (2007) Efficient stochastic sensitivity analysis of discrete event systems. *J. Comput. Phys.*, **221**, 724–738.

Rathinam,M. *et al.* (2010) Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *J. Chem. Phys.*, **132**, 034103.

Sheppard,P. *et al.* (2012) A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *J. Chem. Phys.*, **136**, 34115.