Systems biology

StochSens—matlab package for sensitivity analysis of stochastic chemical systems

Michał Komorowski*, Justina Žurauskienė and Michael P.H. Stumpf*

Division of Molecular Biosciences, Imperial College London, London, UK

Associate Editor: Trey Ideker

ABSTRACT

Motivation: The growing interest in the role of stochasticity in biochemical systems drives the demand for tools to analyse stochastic dynamical models of chemical reactions. One powerful tool to elucidate performance of dynamical systems is sensitivity analysis. Traditionally, however, the concept of sensitivity has mainly been applied to deterministic systems, and the difficulty to generalize these concepts for stochastic systems results from necessity of extensive Monte Carlo simulations.

Results: Here we present a Matlab package, StochSens, that implements sensitivity analysis for stochastic chemical systems using the concept of the Fisher Information Matrix (FIM). It uses the linear noise approximation to represent the FIM in terms of solutions of ordinary differential equations. This is the first computational tool that allows for quick computation of the Information Matrix for stochastic systems without the need for Monte Carlo simulations.

Availability: http://www.theosysbio.bio.ic.ac.uk/resources/stns

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

Contact: mkomor@ippt.gov.pl; M.Stumpf@imperial.ac.uk

1 INTRODUCTION

In the past decade, dynamic modelling of biochemical systems using stochastic approaches has become increasingly popular as the role of stochastic effects has turned out to be significant for the functioning of living cells. The dynamical behaviour of such systems can not be understand by intuition alone but requires computational tools for the analysis. In the case of deterministic models sensitivity analysis is a concept that greatly helps to understand the dependance between parameter values and the dynamic outputs. Various software packages have become available to address this problem [e.g. Maiwald and Timmer (2008); Rodriguez-Fernandez and Banga (2010); D.A. Rand (2008)]. The classical sensitivity coefficient can simply be understood as a derivative of an output Q with respect to a parameter θ. The behaviour of a stochastic system, however, is not defined by an observable Q that can be measured in a reproducible way. Instead it is defined by a distribution with density ψ from which measurements are taken. Therefore, there is a qualitative difference how outputs of stochastic and deterministic systems need

*To whom correspondence should be addressed.

to be quantified. In order to predict a change in the behaviour of a deterministic system resulting from a perturbation in a parameter it is possible to calculate the derivative. In case of a stochastic system one needs to see how a distribution changes. Changes in the distribution ψ with respect to a change in the parameter θ are traditionally quantified as the expected value of a squared infinitesimal change in the logarithm of the density,

$$I(\theta) = E\left(\frac{\partial \log \psi(X, \theta)}{\partial \theta}\right)^2. \tag{1}$$

This quantity is called the Fisher Information (FI) and requires the value of the parameter θ to be known (or at least a plausible range). When calculated for a vector of parameters it is called Fisher Information Matrix (FIM). FI behaves locally as the Kullback–Leibler divergence and is directly related to the curvature of the expected likelihood function, therefore it is a natural measure of local sensitivity [see Supplementary User Manual (UM)]. The possibility to evaluate Equation (1) for stochastic models of chemical kinetics has been previously achieved by the means of Monte Carlo simulation of model variables [see e.g. Gunawan *et al.* (2005); Plyasunov and Arkin (2007)]. Recently we proposed to evaluate Equation (1) by approximating ψ through the linear noise approximation, and expressing the first two momements of ψ and their derivatives in terms of ordinary differential equations (ODEs) [Komorowski *et al.* (2011)].

The applicability of FI extends much further than only to sensitivity analysis. It is a primary tool to study robustness, understood as stability of behaviour under simultaneous changes to model parameters; to guide optimal experimental design; and for identifiability analysis that determines whether certain parameters can be estimated from experimental data of a given type.

Here we present a Matlab package, StochSens, which implements symbolic and numerical procedures to calculate the FIM for stochastic chemical kinetics models as proposed in [Komorowski *et al.* (2011)]. It is the first package that allows computation of the FIM without Monte Carlo simulations; calculations are therefore very efficient and can easily be repeated for different parameter sets and alternative versions of a model corresponding to different experimental setups. The algorithm for FIM computation is accompanied by functions for its analysis, visualization and comparison. Models of Lotka–Voltera systems and Michaelis–Menten kinetics have been used to demonstrate the applicability of the package in the Supplementary Material. Two other models (a model of single gene expression and p53 system) have been also included as examples.

2 METHODS

The package is implemented as a set of Matlab functions. Any model to be analysed using these functions needs to be defined in terms of a stoichiometry matrix, here a Matlab function containing reaction rates, and a vector of parameter values. From these files a set of ODEs is generated using the Matlab Symbolic Toolbox. Equations are then solved using the Matlab ODE solver and the solutions are used to calculate the FIM. The FIM can be calculated for three types of data: time series (TS), time-point (TP) and deterministic (DT). For TS measurements are assumed to be taked at different times from a single trajectory (e.g. following the same cell using fluorescent microscopy). TP measurements at each time point are taken from different trajectories (e.g. end time points of trajectories following different cells). These data reflect experimental setups where the sample is sacrificed to make a measurement (e.g. flow-cytometry). DT data are defined as a solution of deterministic equations with normally distributed measurement error and refer to measurements averaged over population of cells. Graphical output has been implemented to visualize and compare FIMs between data types, parameter sets and perturbations of initial conditions.

2.1 Applicability

The applicability of the package results directly from a special role that the FI plays in the model analysis [Komorowski *et al.* (2011)]. In particular, the package can provide insight resulting from:

(i) *Sensitivity analysis.* The sensitivity coefficients can be directly extracted from the FIM and the most sensitive parameters can be ranked providing information about the origins of the observed stochastic dynamics. In addition, comparison of sensitivity coefficients between stochastic and deterministic version of a model can determine parameters controlling stochasticity: parameters with high sensitivities for a stochastic model and low sensitivities for the corresponding deterministic model control stochastic effects in a system. A set of provided Matlab functions can calculate and analyse sensitivity coefficients.

(ii) *Robustness* of a system is usually understood as its insensitivity to perturbations caused by external conditions. It is usually examined by considering simultaneous perturbations of all model parameters. The FI is a natural tool to quantify changes in systems behaviour in response to perturbations in parameters. For this purpose, it can be visualised as a corresponding quadratic form. Contour lines of these quadratic forms describe sets of points in parameter space that result in a similar change in the behaviour compared to an initial parameter set. Therefore, regions within these ellipsoids are called Neural Spaces (NS); we provide a set of functions to plot and compare NS's across models.

(iii) *Identifiability.* The number of identifiable model parameters is equal to the number of (effectively) non-zero eigenvalues of the FIM. Analysis of eigenvalues therefore provides information as to which parameters can be estimated from a particular data type. Having calculated the FIM, identifiability analysis is straightforward.

(iv) *Optimal experimental design.* The FIM is a key tool to quantify information from particular experimental data. It provides a rationale as to which variables can be measured experimentally, or what perturbation should be applied, or what sampling frequency should be used. Different experimental settings can be encoded within the package and resulting information can be compared to guide experimental design.

2.2 Biological relevance—example

The differences in the FIMs that exist between stochastic and deterministic models is an interesting example highlighting the biological relevance of our package. In order to demonstrate this we use the example of Michaelis–Menten (MM) enzyme kinetics that can be represented as
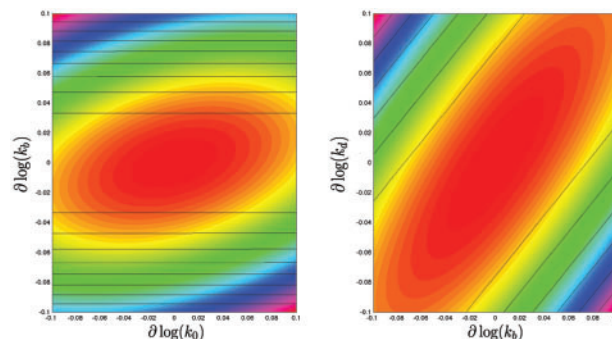

Fig. 1. Neutral spaces for stochastic (heat map) and deterministic (contour plot) versions of Michaelis–Menten enzyme kinetics models for logs of two parameter pairs (k_0, k_b) (left) and (k_b, k_d) (right).

$S + E \xrightleftharpoons[k_1]{k_0} C \xrightarrow{k_2} E + P$. Substrate molecules (S) reversibly bind to an enzyme molecule (E) with forward rate constant k_0 and backward rate constant k_1 to form a complex (C), which then falls apart into the enzyme and a product (P). In addition we assume that (S) molecules are being generated at rate k_b, and product molecules (P) degrade at rate k_d, and that the system has reached its unique steady state. In Fig. 1 we plot neutral spaces that correspond to the FIM for two pairs of parameters (k_0, k_b) and (k_b, k_d) for a stochastic model (heat map) and deterministic model (contour plot). Plotted differences have a clear biological interpretation: in a deterministic regime, in order to keep dynamic behaviour unchanged, change in the arrival rate (k_b) can not be compensated by an increase in a complex formation rate (k_0) as both parameters are approximately orthogonal; in a stochastic regime such a compensation is possible: change in k_b can be to a certain extend compensated by change in k_0. Therefore, if the output of the MM system is considered as a distribution of stochastic trajectories rather than a deterministic mean then such an output is more robust as depicted by the ellipsoids in Fig. 1. Analogous analysis can be used to study the relation between arrival rate k_b and decay rate k_d.

3 DISCUSSION

StochSens is a new, computationally efficient and extendable Matlab package for computation and analysis of FIM's for stochastic (and deterministic) dynamical models of chemical reactions. Compared to sensitivity packages for deterministic models it accounts for the information that is contained in the noise structure. This together with the high interest in stochasticity in biochemical systems imply that our package can find wide applicability and be a basis for further extensions.

REFERENCES

Gunawan,R. *et al.* (2005) Sensitivity analysis of discrete stochastic systems. *Biophysical J.*, **88**, 2530–2540.

Komorowski,M. *et al.* (2011) Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc. Natl Acad. Sci. USA*, **108**, 8645.

Maiwald,T. and Timmer,J. (2008) Dynamical modeling and multi-experiment fitting with potterswheel. *Bioinformatics*, **24**, 2037.

Rand,D.A. (2008) Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation Law. *J R Soc Interface*, **5**, S59–S69.

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

Rodriguez-Fernandez,M. and Banga,J.R. (2010) SensSB: a software toolbox for the development and sensitivity analysis of systems biology models. *Bioinformatics*, **26**, 1675—1676.