Sensitivity Analysis in Biological Modelling

Qihua Huang

October 18, 2012

Outline

1. What is sensitivity analysis?

Outline

1. What is sensitivity analysis?
2. Local sensitivity analysis

Outline

1. What is sensitivity analysis?
2. Local sensitivity analysis
3. Global sensitivity analysis

What is Sensitivity Analysis

Definition(by Nestorov, 1999):

Sensitivity analysis is the systematic investigation of the model response to either

1) perturbation of the model ***quantitative*** factors
(e.g. input and/or parameters)

or

2) variations in the model ***qualitative*** factors
(e.g. structure, connectivity modules or submodels)

Classes of Methods

Local methods: inputs are varied one at a time by a small amount around some fixed point and the effect of individual perturbations on the output are calculated.

Global methods: all inputs are varied simultaneously over their entire input space, typically using a sampling based approach, and the effects on the output of both individual inputs and interactions between inputs are assessed.

Single Parameter Sensitivity

$$S = \frac{(R_a - R_n)/R_n}{(P_a - P_n)/P_n},$$

where R_a and R_n are model responses for altered and nominal parameters, respectively, P_a and P_n are the altered and nominal parameters, respectively.

Sensitivity Matrix

For a general ODE model of the form:

$$\frac{d\mathbf{y}}{dt} = f(\mathbf{y}, \mathbf{p}, t), \quad \mathbf{y}(0) = \mathbf{y}^0,$$

where **y** is the vector of variables, **p** is the m-vector of system parameters. Then

$$y_i(t, \mathbf{p} + \Delta\mathbf{p}) = y_i(t, \mathbf{p}) + \sum_{j=1}^{m} \frac{\partial y_i}{\partial p_j} \Delta p_j + \frac{1}{2} \sum_{l=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 y_i}{\partial p_l p_j} \Delta p_l \Delta p_j + \cdots$$

First-order local sensitivity coefficients: $\dfrac{\partial y_i}{\partial p_j}$.

Sensitivity matrix: $\mathbf{S}(t) = \{s_{ij}\} = \{\dfrac{\partial y_i}{\partial p_j}\}$.

Finite Difference Method (Indirect Method)

The sensitivities can be calculated using a forward difference approximation:

$$s_{ij}(t) \approx \frac{y_i(p_j + \Delta p_j, t) - y_i(p_j, t)}{\Delta p_j}.$$

The indirect method requires at least $m + 1$ runs of the model.

Challenge: the selection of the parameter step size.

What if the step size is too large?

What if the step size is too small?

Direct Differential Method

Absolute sensitivity differential equations:

$$\frac{d}{dt}\frac{\partial \mathbf{y}}{\partial p_j} = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial p_j} + \frac{\partial f}{\partial p_j} = \mathbf{J} \cdot \mathbf{S}_j + \mathbf{F}_j,$$

where

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_k} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_k}{\partial y_1} & \frac{\partial f_k}{\partial y_2} & \cdots & \frac{\partial f_k}{\partial y_k} \end{pmatrix}, \mathbf{F}_j = \begin{pmatrix} \frac{\partial f_1}{\partial p_j} \\ \frac{\partial f_2}{\partial p_j} \\ \cdots \\ \frac{\partial f_k}{\partial p_j} \end{pmatrix}, \mathbf{S}_j = \frac{\partial \mathbf{y}}{\partial p_j} = \begin{pmatrix} s_{1,j} \\ s_{2,j} \\ \cdots \\ s_{k,j} \end{pmatrix}$$

Letting $s_{i,j}^0 = \delta(P_j - x_i^0)$ and solving

$$\begin{cases} \dfrac{d\mathbf{y}}{dt} = f(\mathbf{y}, \mathbf{p}, t), \quad \mathbf{y}(0) = \mathbf{y}^0, \\ \dfrac{d\mathbf{S}_j}{dt} = \mathbf{J} \cdot \mathbf{S}_j + \mathbf{F}_j, \quad \mathbf{S}_j(0) = \mathbf{S}_0 \end{cases}$$

both $\mathbf{y}(t)$ and $\dfrac{\partial \mathbf{y}}{\partial p_j}$ can be determined simultaneously.

In reality, the following relative sensitivities are normally used instead of s_{ij}.

$$\bar{s}_{i,j} = \frac{\partial y_i / y_i}{\partial p_j / p_j} = \frac{\partial y_i}{\partial p_j} \cdot \frac{p_j}{y_i}$$

Feature Sensitivity Analysis

In many cases we should be more interested in the sensitivity of ***aspects*** of the model output rather than the sensitivity of the output at a given time point.

Questions:

What influences the maximum value of the species concentration?

How does the period of an oscillatory solution vary with the model parameters?

Feature sensitivities can also be derived from so called “elementary sensitivities" calculated via the direct method. The perturbed solution can be approximated as:

$$\tilde{\mathbf{y}}_j \approx \mathbf{y} + \mathbf{S}_j \Delta p_j,$$

where $\mathbf{S}_j$ are the sensitivities of the output to parameter p_j.

The feature of interest can now be evaluated from the original and approximated perturbed solution and its sensitivity to p_j calculated as:

$$S_{F,j} = \frac{\tilde{F}_j - F}{\Delta p_j}.$$

Limitations of Local SA

1. Only investigate the behavior of a model in the immediate region around the nominal parameter values. In biology, input values are often very uncertain and cover large ranges which can not be investigated using local techniques.

2. Local techniques only consider changes to one parameter at a time, with all other parameters fixed to their nominal values. In biological systems it is likely that interactions between parameters will be important.

Sampling Based Methods

Use Monte-Carlo (MC) techniques to explore the mapping between uncertain model parameters and outputs.

For a model with k parameters $\mathbf{p} = [p_1, p_2, \cdots, p_k]$ a general sampling based approach involves five main steps:

Step 1. Define distributions $D_1, D_2, \cdots, D_k$ that characterize the uncertainties in the parameters $\mathbf{p}$

- Parametric fitting to known distributions
- Using non-parameter density estimation techniques
- Uniform distribution

Step 2. Generates a sample of size N, $X_1, X_2, \cdots, X_N$, from the distributions defined in step 1

- Random sampling
- Latin hypercube sampling

Step 3. Evaluate the model for each element in the input sample to obtain a set of model outputs, $\mathbf{y}(\mathbf{p}_i)$, $i = 1, 2, \cdots, N$

- Model and application

Step 4. Quantify and display the uncertainty in the model outputs

- Mean value and variance
- Plotting the PDF or CDF

Step 5. Explore the mapping between uncertain inputs and the output uncertainty

- Examine scatter plots of the output against parameter values
- Regression or correlation analysis
- Partial rank correlation coefficients

Reference:"*Sensitivity and uncertainty analysis of complex model of disease transmission: an HIV Model, as an Example*" (by S.M. Blower and H. Dowlatabadi)

Variance Based Methods

For a general model of the form:

$$Y = f(\mathbf{X}),$$

where $\mathbf{X} = (x_1, x_2, \cdots, x_k)$ is a k-vector of uncertain model factors.

If all the factors are allowed to vary over their entire range of values, then the uncertainty in the model output can be quantified by its unconditional variance $V_{\mathbf{X}}(Y)$.

Question: How does removing the uncertainty in factor x_i reduce the variance in the model output Y?

The effect on the variance of fixing is given by the conditional variance:

$$V_{\mathbf{X}_{-i}}(Y|x_i = x_i^*).$$

The true value of x_i^* is not known so the conditional variance is averaged over all possible values of x_i:

$$E_{x_i}(V_{\mathbf{X}_{-i}}(Y|x_i)) := E(V(Y|x_i))$$

Question: What is the relationship between the value of $E(V(Y|x_i))$ and the importance of factor x_i?

The smaller the value the larger the influence of on the model output and the more important the factor.

The total variance is a constant and can be expressed using the "*law of total variance*" as:

$$V(y) = V(E(Y|x_i)) + E(V(Y|x_i)).$$

Selecting factors with small as important is equivalent to selecting those with high values of $V_i = V(E(Y|x_i))$.

The method of Sobol

The function $f(\mathbf{X}) = f(x_1, x_2, \cdots, x_k)$ under investigation is defined in the $k-$ dimensional unit cube: $\Omega^k = \{\mathbf{X} | 0 \leq x_i \leq 1; i = 1, \cdots, n\}$, and can be written as the sum:

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_{1 \leq i < j \leq k} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\cdots,k}(x_1, \cdots, x_k) \quad (2.1)$$

provided that f_0 is a constant and the integral of every term over any of its variables is zero. Then

$$f_0 = \int_{\Omega^k} f(\mathbf{X}) d\mathbf{X}.$$

The total variance of f(X) can be written as:

$$V = \int_{\Omega^k} f^2(\mathbf{X}) d\mathbf{X} - f_0^2.$$

The method of Sobol

This can also be decomposed in the same manner as the function itself:

$$V = \sum_{i=1}^{k} V_i + \sum_{1 \le i < j \le k} V_{ij} + \cdots + V_{1,2,\cdots,k}. \tag{2.2}$$

The terms of this decomposition are the contributions to the variance from term $f_{i_1,\cdots,i_s}$ in (2.1) and are given by:

$$V_{i_1,\cdots,i_s} = \int_0^1 \cdots \int_0^1 f^2_{i_1,\cdots,i_s}(x_{i_1}, \cdots, x_{i_s}) dx_{i_1} \cdots dx_{i_s}.$$

The Sobol indices are then defined as:

$$S_{i_1,\cdots,i_s} = \frac{V_{i_1,\cdots,i_s}}{V}.$$

The method of Sobol

The key to the Sobol method is that the integrals can be evaluated using Monte Carlo integrals.

$$\hat{f}_0 = \frac{1}{N} \sum_{m=1}^{N} f(\mathbf{X}_m)$$

$$\hat{V} = \frac{1}{N} \sum_{m=1}^{N} f^2(\mathbf{X}_m) - \hat{f}_0^2$$

The first order effects require estimates for the $V_i s$ which are given by

$$\hat{V}_i = \frac{1}{N} \sum_{m=1}^{N} f(\mathbf{X}_1) f(\mathbf{X}_{m(\sim i)}, \mathbf{X}_{1,i}) - \hat{f}_0^2,$$

we “resample" all factors except the factor of interest x_i.

The method of Sobol

The second order effects require estimates for V_{ij} which are given by

$$\hat{V}_{ij} = \frac{1}{N}\sum_{m=1}^{N} f(\mathbf{X}_1) f(\mathbf{X}_{m(\sim ij)}, \mathbf{X}_{1,i}, \mathbf{X}_{1,j}) - \hat{f}_0^2 - \hat{V}_i - \hat{V}_j,$$

where all factors except x_i and x_j are resampled in the second term in the product.

Similar expressions can be derived for the higher order terms.

Each effect requires the evaluation of the model for an additional sample of size N. The decomposition in equation (2.2) contains $2^k - 1$ terms, therefore the total cost of evaluating all effects is $N2^k$.

Summarization

1. Definition of Sensitivity analysis
2. Local sensitivity analysis
    - Basic formula for single parameter sensitivity analysis
    - Sensitivity matrix
    - Finite difference method
    - Direct differential method
    - Feature sensitivity analysis
3. Global sensitivity analysis
    - Sampling based method
    - Variance based method
    - The method of Sobol