Reliability Engineering and System Safety 110 (2013) 60–67

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Monte Carlo simulation for moment-independent sensitivity analysis

Pengfei Wei, Zhenzhou Lu*, Xiukai Yuan

Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

ARTICLE INFO

Article history:
Received 9 May 2012
Received in revised form
4 September 2012
Accepted 18 September 2012
Available online 27 September 2012

Keywords:
Monte Carlo simulation
Moment-independent sensitivity analysis
Delta indices
Kernel density estimation

ABSTRACT

The moment-independent sensitivity analysis (SA) is one of the most popular SA techniques. It aims at measuring the contribution of input variable(s) to the probability density function (PDF) of model output. However, compared with the variance-based one, robust and efficient methods are less available for computing the moment-independent SA indices (also called delta indices). In this paper, the Monte Carlo simulation (MCS) methods for moment-independent SA are investigated. A double-loop MCS method, which has the advantages of high accuracy and easy programming, is firstly developed. Then, to reduce the computational cost, a single-loop MCS method is proposed. The later method has several advantages. First, only a set of samples is needed for computing all the indices, thus it can overcome the problem of "curse of dimensionality". Second, it is suitable for problems with dependent inputs. Third, it is purely based on model output evaluation and density estimation, thus can be used for model with high order (>2) interactions. At last, several numerical examples are introduced to demonstrate the advantages of the proposed methods.

1. Introduction

Sensitivity analysis (SA) is a study of how "uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [1]. It is a very useful tool for model simplification, importance ranking, risk reduction and other purposes. During the past few decades, a variety of SA techniques have been proposed by researchers for different purposes [2–9]. Among these techniques, the variance-based one developed by Sobol [3] and Homma and Saltelli [4] and the moment-independent one developed by Borgonovo [9] are the most popular.

The variance-based SA aims at distributing the model output variance to different sets of model inputs by looking at the entire distribution ranges of those inputs. It is global, quantitative and model free, thus has been studied widely by practitioners in the past years. Nowadays, there have been a lot of smart methods available for computing the variance-based SA indices [10–13].

The moment-independent SA focuses on finding those inputs that, if fixed at their distribution ranges, will lead to the greatest shift in the probability density function (PDF) of model output on average. It is also global, quantitative and model free, and additionally, it is moment-independent, thus attracts more and more attentions of practitioners recently. During the past few years, some efforts have been devoted to developing efficient and robust methods for computing the moment-independent SA indices (also called delta indices). In the original paper, Borgonovo proposed the PDF-based method for computing the delta indices [9]. This is a computationally expensive double-loop simulation method, and also the precision of estimates suffers from calculating the intersection points of the unconditional and conditional PDF of the model output. To improve the accuracy of the estimates, Liu developed the CDF-based method, however, as pointed by Liu "*for a computationally intensive model, when the total computational time is mainly due to the time of running the model, the improvement of the computational efficiency by the CDF-based method can be negligible*" [14]. In another paper, Borgonovo proposed the emulation method for expressing the input–output relationship by metamodel, and then computing the delta indices based on the metamodel [15]. Both the State Dependent Parameter (SDP) [12,16,17] and kriging emulator are investigated. This method can drastically reduce the computational cost, thus receives more and more attentions in the area of SA. The estimate precision of these methods mainly depends on the metamodel. If the input–output mapping contains high order interactions, the metamodel may often fail to capture the structure feature of the model. In Ref. [18], Castaings used the quadrature method to deal with the one dimensional integral given in the definition of delta indices for individual input. Compared with the PDF-based and CDF-based methods, this method is computationally more efficient. However, it is still a double-loop method, and the computational cost increases with the number input variables. Despite these works by different researchers, compared with the variance-based SA indices, robust and efficient methods are less available for computing the delta indices.

* Corresponding author. Tel./fax: +86 29 88460480.
E-mail address: zhenzhoulu@nwpu.edu.cn (Z. Lu).

0951-8320/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2012.09.005

In this paper, we firstly propose a double-loop Monte Carlo simulation (MCS) method for computing the delta indices. This method is purely based on model evaluation and univariate density estimation, and it has the advantages of high-accuracy and usability, but its computational cost is unacceptable for engineering models. Then, to substantially improve the computational efficiency, a single-loop MCS method is developed. This single-loop MCS method has several advantages. First, only a set of samples are needed for calculating all the delta indices, thus it is computationally efficient and overcomes the problem of "curse of dimensionality". Second, it is suitable for models with dependent inputs. Thirdly, compared with the emulation method, it is based purely on model evaluation and density estimation, thus can be employed to deal with the problem with high order (>2) interactions terms.

The rest of this paper is organized as follows. Section 2 reviews the original definition of the delta indices. Section 3 proposes the double-loop MCS method and Section 4 develops the single-loop MCS method for computing the delta indices. Section 5 introduces three test examples to demonstrate the advantages of the proposed methods. Section 6 gives conclusions.

2. Review of the moment-independent sensitivity analysis technique

Consider a computational model represented by $Y=g(\boldsymbol{X})$, where $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ is the vector of random input variables, Y is the model output of interest. The joint PDF of $\boldsymbol{X}$ is denoted as $f_X(\boldsymbol{x})$, and the marginal PDF of X_i ($i=1,2,\ldots,n$) can be obtained as following:

$$f_{X_i}(x_i)=\int\ldots\int f_{\mathbf{X}}(\mathbf{x})\prod_{k=1,k\neq i}^{n}\mathrm{d}x_k \tag{1}$$

To measure the effect of the uncertainty of an individual input X_i on the PDF $f_Y(y)$ of model output Y, Borgonovo proposed the following moment-independent SA index (also named as delta index) for X_i [9]:

$$\delta_i=\frac{1}{2}E_{X_i}(s(X_i))=\int s(X_i)f_{X_i}(x_i)\mathrm{d}x_i \tag{2}$$

where $s(X_i)$ is the measure of the shift between the unconditional PDF $f_Y(y)$ and conditional PDF $f_{Y|X_i}(y)$ on X_i of the model output Y, and its expression is:

$$s(X_i)=\int|f_Y(y)-f_{Y|X_i}(y)|\mathrm{d}y \tag{3}$$

Further, Borgonovo defined the moment-independent SA index for a group of inputs $\mathbf{R}=(X_{i_1},X_{i_2},\ldots,X_{i_r})$ as follows:

$$\delta_{i_1,i_2,\ldots,i_r}=\frac{1}{2}E_{\mathbf{R}}(s(\mathbf{R}))=\frac{1}{2}\int f_{X_{i_1},X_{i_2},\ldots,X_{i_r}}(x_{i_1},x_{i_2},\ldots,x_{i_r})\times\left(\int|f_Y(y)-f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)|\mathrm{d}y\right)\mathrm{d}x_{i_1}\mathrm{d}x_{i_2}\ldots\mathrm{d}x_{i_r} \tag{4}$$

where $f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)$ is the PDF of model output Y conditional on $\boldsymbol{R}$, $f_{X_{i_1},X_{i_2},\ldots,X_{i_r}}(x_{i_1},x_{i_2},\ldots,x_{i_r})$ is the joint PDF of $\boldsymbol{R}$, and it can be given by

$$f_{X_{i_1},X_{i_2},\ldots,X_{i_r}}(x_{i_1},x_{i_2},\ldots,x_{i_r})=\int\ldots\int f_{\mathbf{X}}(\mathbf{x})\prod_{k=1,k\neq i_1,i_2,\ldots,i_r}^{n}\mathrm{d}x_k \tag{5}$$

In Ref. [9], Borgonovo derived five properties of δ_i, which are shown in Table 1. Property no.1 indicates that the lower bound of δ_i is zero, and the upper bound is unity. Property no. 2 shows that, if Y is independent of X_i, then δ_i equals to zero. This property is obvious since $f_Y(y)=f_{Y|X_i}(y)$ as Y is independent of X_i. Property no. 3 indicates that delta index of all input variables equals unity. Property no. 4 indicates that if Y is dependent on X_i but independent of X_j, then $\delta_{ij}=\delta_i$, and vice verse.

Let

$$\delta_{j|i}=\frac{1}{2}\int\int\left(\int|f_{Y|X_i}(y)-f_{Y|X_i,X_j}(y)|dy)f_{X_i,X_j}(x_i,x_j\right)\mathrm{d}x_i\mathrm{d}x_j \tag{6}$$

Then from property no. 5, $\delta_i\leq\delta_{ij}\leq\delta_i+\delta_{j|i}$ holds. This property provides bounds for the possible values of δ_{ij}. The geometrical interpretation of this property can be found in Ref. [9]. δ_i can be seen as the average distance measure between $f_Y(y)$ and $f_{Y|X_i}(y)$. Similarly, $\delta_{j|i}$ is that between $f_{Y|X_i}(y)$ and $f_{Y|X_i,X_j}(y)$, δ_{ij} measures the average distance between $f_Y(y)$ and $f_{Y|X_i,X_j}(y)$. $\delta_{ij}\leq\delta_i+\delta_{j|i}$ is nothing but the triangle inequality. If Y is independent of X_i, $\delta_{j|i}=0$, then by property no. 5, $\delta_i=\delta_{ij}$ holds. This conclusion is consistent with property no. 4. In this case, the vector from $f_Y(y)$ to $f_{Y|X_i}(y)$ coincides with the vector from $f_Y(y)$ to $f_{Y|X_i,X_j}(y)$.

Up to now, we have briefly reviewed the delta indices and their properties. In the next section, we introduce the double-loop MCS method for computing δ_i.

3. Double-loop Monte Carlo simulation method

From Eq. (2), we know that

$$\delta_i=\frac{1}{2}E_{X_i}(s(X_i)) \tag{7}$$

where $s(X_i)$ can be derived to be:

$$\begin{aligned}s(X_i)&=\int|f_Y(y)-f_{Y|X_i}(y)|\mathrm{d}y\\&=\int\left|\frac{f_Y(y)}{f_{Y|X_i}(y)}-1\right|f_{Y|X_i}(y)\mathrm{d}y\\&=E_{Y|X_i}\left(\left|\frac{f_Y(y)}{f_{Y|X_i}(y)}-1\right|\right)\end{aligned} \tag{8}$$

where the subscript $Y|X_i$ indicates that the expectation is taken with respect to the conditional PDF $f_{Y|X_i}(y)$.

Eqs. (7) and (8) indicate that δ_i can be expressed in the form of double-loop expectation, thus can be estimated by the double-loop MCS method. Similarly, for a set of input variables $\boldsymbol{R}$, we have:

$$\delta_{i_1,i_2,\ldots,i_r}=\frac{1}{2}E_{\mathbf{R}}(s(\mathbf{R})) \tag{9}$$

where

$$\begin{aligned}s(\mathbf{R})&=\int|f_Y(y)-f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)|\mathrm{d}y\\&=\int\left|\frac{f_Y(y)}{f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)}-1\right|f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)\mathrm{d}y\\&=E_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}\left(\left|\frac{f_Y(y)}{f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)}-1\right|\right)\end{aligned} \tag{10}$$

Eqs. (9) and (10) indicate that $\delta_{i_1,i_2,\ldots,i_r}$ can also be expressed as a double-loop expectation, thus can also be estimated by a double-loop MCS method.

It is shown by Eqs. (7)–(10) that, for computing δ_i by the double-loop MCS method, one needs to estimate the PDF $f_Y(y)$ and $f_{Y|X_i}(y)$, and for computing $\delta_{i_1,i_2,\ldots,i_r}$, one needs to estimate $f_Y(y)$ and $f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)$. Since $f_Y(y)$, $f_{Y|X_i}(y)$ and $f_{Y|X_{i_1},X_{i_2},\ldots,X_{i_r}}(y)$ are univariate density functions, all of them can be easily estimated by the density estimation methods. In this paper, we only consider the calculation of the delta indices of single input variable, i.e., δ_i, thus we only need to estimate $f_Y(y)$ and $f_{Y|X_i}(y)$.

The density estimation methods can be divided into two groups: the parametric one and the nonparametric one. The precondition of using the parametric density estimation (PDE)