Reliability Engineering and System Safety 131 (2014) 66–82

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Space-partition method for the variance-based sensitivity analysis: Optimal partition scheme and comparative study

Qingqing Zhai, Jun Yang*, Yu Zhao

School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form
8 May 2014
Accepted 27 June 2014
Available online 5 July 2014

Keywords:
Global sensitivity analysis
Variance-based sensitivity indices
Sample space partition
Minimized estimation variance
Monte Carlo sampling

ABSTRACT

Variance-based sensitivity analysis has been widely studied and asserted itself among practitioners. Monte Carlo simulation methods are well developed in the calculation of variance-based sensitivity indices but they do not make full use of each model run. Recently, several works mentioned a scatter-plot partitioning method to estimate the variance-based sensitivity indices from given data, where a single bunch of samples is sufficient to estimate all the sensitivity indices. This paper focuses on the space-partition method in the estimation of variance-based sensitivity indices, and its convergence and other performances are investigated. Since the method heavily depends on the partition scheme, the influence of the partition scheme is discussed and the optimal partition scheme is proposed based on the minimized estimator's variance. A decomposition and integration procedure is proposed to improve the estimation quality for higher order sensitivity indices. The proposed space-partition method is compared with the more traditional method and test cases show that it outperforms the traditional one.

1. Introduction

Sensitivity analysis (SA) aims at investigating the sensitivity of the model output with respect to its input parameters. It involves a variety of techniques, which are generally classified into local SA techniques and global SA techniques. In the domain of local SA, the sensitivity of the model output about a certain parameter is studied by varying the particular parameter around its nominal value while keeping other parameters at their nominal states [1–3]. It is often the case, however, that the parameters in the model are not exactly known, which violates the assumption of the local SA. Typically, it is assumed that the model output y is linked to the model inputs $\mathbf{x}=[x_1,x_2,\dots,x_n]^T$ by a deterministic function $g(\cdot)$, i.e. $y=g(\mathbf{x})$, where the model inputs are not known with certainty. Then the uncertainty of model inputs will propagate through $g(\cdot)$ and lead to the output non-deterministic. Studying the influence of the inputs' uncertainty on the output's uncertainty is what global SA concerns.

The scope of global SA is to rank the importance of various sources of uncertainty that result in the uncertainty of the output. A number of global SA techniques are available, such as non-parametric techniques [4–6], meta-model techniques [7–12], variance-based techniques [13–16], and moment independent techniques [17–19]. Among all these different global SA techniques, the variance-based techniques attract most attention and have asserted themselves among practitioners. To calculate the variance-based sensitivity indices, two popular approaches are Fourier Amplitude Sensitivity Test (FAST) [20–23] and Monte Carlo sampling [14–16,24]. In FAST, variance-based sensitivity indices are calculated in terms of the coefficients of the multiple Fourier series expansion of the output function, which is powerful in estimating the "first-order" effect [25,26]. Recently, Xu and Gertner [27] show that FAST can also be used to estimate higher order effects, which further extends its application.

Sampling is always a straightforward way to investigate the response of complex functions and generally effective. However, to estimate the sensitivity indices, a great amount of model runs may be required for the general sampling-based methods. It is easy to handle if the model is simple, but prohibitive for complex models (e.g., a single model run needs several hours or more). To estimate the sensitivity indices with as fewer model runs as possible, many efforts have been devoted to develop efficient algorithms. In this respect, one direction is to fill the input space with proper representative samples; another is to make best use of model runs. To accomplish the former, Latin hypercube sampling (LHS) was proposed in the late 1970s [28] and has been well studied since then. For the latter, different formulas to calculate the variance-based indices have been proposed [14,29–31]. The traditional FAST method by Cukier et al. in 1970s [20] is another good example in best using model samples with the use of cyclic samples. However, almost all of these methods cannot fully exploit the model runs. In fact, most model runs are used only once in the sensitivity indices calculation [14].

* Corresponding author. Tel./fax: +86 10 82316003.
E-mail address: tomyj2001@buaa.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.ress.2014.06.013
0951-8320/

To make best use of the sample runs (and in turn reduce the cost), Plischke et al. [32] proposed a "space-partition" method to estimate the moment-independent importance measure [18]. The space-partition method is universal, which is applicable to any given data and independent of the generation method. Specifically, it can be used to estimate the sensitivity indices of models with independent or dependent inputs. Though they mentioned the extension of the method to estimate other sensitivity indices, such as the variance-based sensitivity indices, they discussed less of the estimator's properties for the variance-based indices (e.g., the convergence of the estimator and the optimal partition scheme). In this paper, we focus on the space-partition approach in the estimation of the variance-based sensitivity indices, including the first order indices and the higher order extension. Since the performance of the space-partition method heavily depends on the partition scheme, the partition scheme for the space-partition estimator is discussed and the optimal partition scheme based on the minimized estimator's variance is proposed.

The remainder of the paper is organized as follows. Section 2 briefly reviews the variance-based sensitivity indices and their estimation by Monte Carlo simulation. Section 3 introduces the space-partition estimation for the first order sensitivity indices and its extension to higher order indices. Section 4 discusses the influence of different partition schemes for the space-partition estimator and proposes an optimal partition scheme based on the minimized estimator's variance. For the higher order sensitivity indices, a decomposition and integration procedure is proposed to further improve the quality of the estimation. Section 5 illustrates the performance of the space-partition method and its comparison with the more traditional substituted-column method. The case with dependent inputs is also studied to illustrate the applicability of the proposed method. Conclusions and discussions are given in the end.

2. Variance-based sensitivity indices

In this section, we briefly review the variance-based sensitivity indices and their estimation by the Monte Carlo sampling method.

2.1. Variance-based sensitivity indices

In the context of variance-based sensitivity indices, people often consider such a model $y=g(\mathbf{x})$, where $\mathbf{x}=[x_1,x_2,\ldots,x_n]^T$ is the model inputs with uncertainty and $g(\cdot)$ is a scalar, deterministic function. The uncertainty of $\mathbf{x}$ is often modeled in a probabilistic framework, i.e., $\mathbf{x}$ is assumed to be a random vector with joint probability density function (PDF) $f_{\mathbf{x}}(x)$. Let us denote the random input vector as $\mathbf{X}=[X_1,X_2,\ldots,X_n]^T$. The uncertainty of $\mathbf{x}$ propagates through $g(\cdot)$ and results in that y is also a random variable, i.e., Y. The uncertainty of the output Y is often characterized by its variance $V(Y)$. To investigate the influence of a certain input X_i on the output, it is intuitive to assume the true value of X_i is known and quantify the variance reduction in the output, i.e., $V(Y)-V(Y|X_i=x_i^0)$, where x_i^0 is the true value of X_i. Since the true value is never known, one can utilize $V(Y)-E_{X_i}(V(Y|X_i))$ to represent the expected variance reduction in the output, i.e., the sensitivity of Y about X_i.

With the well-known equality

$$V(Y)=V_{X_i}(E(Y|X_i))+E_{X_i}(V(Y|X_i)), \tag{1}$$

the first-order effect of X_i is defined as follows after normalization:

$$S_i=\frac{V(Y)-E_{X_i}(V(Y|X_i))}{V(Y)}=1-\frac{E_{X_i}(V(Y|X_i))}{V(Y)}=\frac{V_{X_i}(E(Y|X_i))}{V(Y)}, \tag{2}$$

By definition, one can note that $0\le S_i\le 1$. In a similar way, the joint effect of X_i and X_j can be defined as

$$S_{\{i,j\}}=\frac{V_{X_i,X_j}(E(Y|X_i,X_j))}{V(Y)}. \tag{3}$$

Higher order sensitivity indices can be defined similarly. Another popular variance-based measure is the total order effect [29]:

$$S_{T_i}=\frac{E_{\mathbf{X}_{\sim i}}(V(Y|\mathbf{X}_{\sim i}))}{V(Y)}=1-\frac{V_{\mathbf{X}_{\sim i}}(E(Y|\mathbf{X}_{\sim i}))}{V(Y)} \tag{4}$$

where $\mathbf{X}_{\sim i}=\{X_1,X_2,\ldots,X_{i-1},X_{i+1},\ldots,X_n\}$. The index can be interpreted as the expected remaining uncertainty in the output when all the inputs except X_i are fixed.

As can be noticed, the key in the calculation of the variance-based sensitivity indices lies in calculating the expectation of the conditional variance of Y, or the variance of the conditional expectation of Y.

2.2. Variance-based sensitivity indices estimation

Consider the estimation of the first-order sensitivity index S_i by Monte Carlo sampling. A crude sampling estimation can be accomplished in the following steps.

Step 1: Sample from the joint PDF $f_{\mathbf{X}}(\mathbf{x})$. Evaluate the model output with these samples and estimate the variance of Y by the sample variance $\hat{V}(Y)$.
Step 2: Sample for the random variable X_i with respect to its marginal PDF $f_{X_i}(x)$. Denote these samples as $\{x_i^1,\ldots,x_i^m\}$.
Step 3: For each x_i^j $(1\le j\le m)$, sample according to the conditional PDF $f_{\mathbf{X}|X_i=x_i^j}(\mathbf{x}_{\sim i})$. Estimate the conditional variance of $(Y|X_i=x_i^j)$ after evaluating the model output with these conditional samples.
Step 4: Estimate the expectation of the conditional variance $E_{X_i}(V(Y|X_i))$ by the arithmetic average of $\hat{V}(Y|X_i=x_i^j)$, i.e.

$$\hat{E}_{X_i}(V(Y|X_i))=\frac{1}{m}\sum_{j=1}^{m}\hat{V}(Y|X_i=x_i^j). \tag{5}$$

Then S_i can be estimated by

$$\hat{S}_i=1-\frac{\hat{E}_{X_i}(V(Y|X_i))}{\hat{V}(Y)}. \tag{6}$$

A double-loop algorithm needs to be performed to estimate all the first order sensitivity indices. Higher order sensitivity indices can also be estimated in a similar manner. It is apparent that each sample is used only once in the estimation and the information is greatly wasted. To exploit the samples, the substituted-column method is proposed when the inputs are independent [14,33]. In this method, one first independently samples two $m\times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, where m denotes the number of samples and n is the dimension of $\mathbf{X}$. Then the matrix $\mathbf{B}_{\mathbf{A}}^{(i)}$ is constructed by substituting the ith column of $\mathbf{B}$ with the ith column of $\mathbf{A}$. The variance of the conditional expectation of Y is then estimated by

$$\hat{V}_{X_i}(E(Y|X_i))=\frac{1}{m}(g(\mathbf{A}))^T(g(\mathbf{B}_{\mathbf{A}}^{(i)})-g(\mathbf{B})), \tag{7}$$

where $g(\mathbf{A})$ is an $m\times 1$ vector whose elements correspond to the rows of $\mathbf{A}$, and the superscript "T" indicates transpose of matrices. With Eq. (7), one requires $m\times(n+2)$ different samples and thus $m\times(n+2)$ model runs to estimate all the n first order sensitivity indices. However, $g(\mathbf{B}_{\mathbf{A}}^{(i)})$ is only used in the estimation of S_i, yet it also contains information about other inputs. The second order sensitivity indices can also be estimated by the