Reliability Engineering and System Safety 107 (2012) 107–114

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

A variance-based sensitivity index function for factor prioritization

Douglas L. Allaire *, Karen E. Willcox

Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

ARTICLE INFO

Article history:
Received 15 October 2010
Received in revised form
15 August 2011
Accepted 22 August 2011
Available online 3 September 2011

Keywords:
Global sensitivity analysis
Factor prioritization
Variance reduction
Rejection sampling

ABSTRACT

Among the many uses for sensitivity analysis is factor prioritization—that is, the determination of which factor, once fixed to its true value, on average leads to the greatest reduction in the variance of an output. A key assumption is that a given factor can, through further research, be fixed to some point on its domain. In general, this is an optimistic assumption, which can lead to inappropriate resource allocation. This research develops an original method that apportions output variance as a function of the amount of variance reduction that can be achieved for a particular factor. This variance-based sensitivity index function provides a main effect sensitivity index for a given factor as a function of the amount of variance of that factor that can be reduced. An aggregate measure of which factors would on average cause the greatest reduction in output variance given future research is also defined and assumes the portion of a particular factors variance that can be reduced is a random variable. An average main effect sensitivity index is then calculated by taking the mean of the variance-based sensitivity index function. A key aspect of the method is that the analysis is performed directly on the samples that were generated during a global sensitivity analysis using rejection sampling. The method is demonstrated on the Ishigami function and an additive function, where the rankings for future research are shown to be different than those of a traditional global sensitivity analysis.

1. Introduction

Sensitivity analysis of model output has been defined as the determination of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model factors [1]. Sensitivity analysis defined in this manner is often referred to as *global sensitivity analysis*, owing to the fact that entire factor distributions are considered in the apportionment process. Since what is meant by the term "uncertainty" is typically case dependent, several indicators have been developed to apportion different measures of uncertainty among model factors. These indicators are often based on screening methods [2], variance-based methods [3–6], entropy-based methods [6,7], non-parametric methods [6,8], and moment-independent approaches [9–11]. This paper focuses on the development and demonstration of an extension of traditional variance-based global sensitivity analysis that considers the change in output variance caused by a change in factor variance that may arise from researching a factor further. In general, variance-based global sensitivity analysis is the standard practice for determining how each factor contributes to output uncertainty when output variance is considered sufficient to describe output variability [4,5].

Variance-based global sensitivity analysis is a rigorous method for apportioning output variance [3,12]. The method has been applied in a wide variety of applications including hydraulic modeling [13], aviation environmental modeling [14], nuclear waste disposal [4], robust mechanical design practices [15], and many others. The two main metrics computed in variance-based global sensitivity analysis are the main effect sensitivity indices proposed by Sobol' [16] and the total effect sensitivity indices proposed by Homma and Saltelli [5]. One of the primary uses of global sensitivity analysis is in the context of *factor prioritization* [3]. In this setting, the objective is to determine which factor, on average, once fixed to its true value, will lead to the greatest reduction in output variance. It has been established by Saltelli et al. [3] and Oakley and O'Hagan [17] that the main effect sensitivity indices are appropriate measures for ranking factors in this setting, however, as noted in Oakley and O'Hagan [17], it is rarely possible to learn the true value of any uncertain factor, and thus these sensitivity indices only suggest the potential for reducing uncertainty in an output through new research on a factor. Given that it is rarely possible to obtain the true value of any uncertain factor, the assumption that a given factor will be fixed to some point on its domain is a major limitation in the use of main effect sensitivity indices for use in allocating resources aimed at reducing output variance.

To account for the inherent limitations in using global sensitivity analysis results for directing future research, a new method

* Corresponding author.
E-mail address: dallaire@mit.edu (D.L. Allaire).

0951-8320/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2011.08.007

that apportions output variance as a function of the amount of variance reduction that can be achieved for a particular factor has been developed. This function is called the variance-based sensitivity index function. By assuming the portion of a particular factor's variance that can be reduced is a random variable, the mean of this function can be taken to provide average main effect sensitivity indices for ranking purposes. A key aspect of the method is that the analysis is performed directly on the factor and output samples that were generated during a global sensitivity analysis using a rejection sampling technique for Monte Carlo simulation proposed in Beckman and McKay [18]. The derivation of the method is given in Section 2, which is followed by a demonstration of the method on the well-known Ishigami function [19] and a purely additive function in Section 3. Conclusions are presented in Section 4.

2. Methodology

In the following subsections, the method is derived using main effect sensitivity indices from global sensitivity analysis. The notion of a reasonable distribution that could arise through future research on a given factor is considered as well as the rejection sampling technique and how it can be employed to reuse model evaluations from a Monte Carlo simulation.

2.1. Derivation

Consider a generic model, $Y=f(\mathbf{x})$, where $\mathbf{x}=[X_1,\ldots,X_k]^T$, and $X_1,\ldots,X_k$ are random variables on the measurable space $(\mathbb{R},\mathcal{B})$, and $f:\mathbb{R}^k\rightarrow\mathbb{R}$ is $(\mathcal{B}^k,\mathcal{B})$-measurable. Then Y is a random variable, and by definition, the variance of Y can be decomposed according to

$$\mathrm{var}(Y)=\mathbb{E}[\mathrm{var}(Y|X_i)]+\mathrm{var}(\mathbb{E}[Y|X_i]), \tag{1}$$

for any X_i, where $i\in\{1,\ldots,k\}$. According to Saltelli et al. [3], the goal of factor prioritization is the identification of which factor, once fixed at its true value, would reduce the variance of Y the most. Since it is not known *a priori* a given factor's true value, factor prioritization is carried out by identifying the factors which, on average, once fixed, would cause the greatest reduction in the variance of Y. The average amount of variance remaining once a given factor is fixed is just $\mathbb{E}[\mathrm{var}(Y|X_i)]$ for any factor X_i. Thus, according to Eq. (1), the average amount of the variance of Y that could be reduced through fixing factor X_i somewhere on its domain is $\mathrm{var}(\mathbb{E}[Y|X_i])$. Global sensitivity analysis uses this fact for factor prioritization by considering main effect sensitivity indices, which take the form

$$S_i=\frac{\mathrm{var}(\mathbb{E}[Y|X_i])}{\mathrm{var}(Y)}, \tag{2}$$

where S_i is the main effect sensitivity index of factor X_i. The main effect sensitivity index can then be used as a measure of the proportion of the variance of Y that is expected to be reduced once factor X_i is fixed to its true value.

The calculation of main effect sensitivity indices in a global sensitivity analysis is most commonly done using either the Fourier Amplitude Sensitivity Test (FAST) method or the Sobol' method [5,12,20,21]. The FAST method is based on Fourier transforms, while the Sobol' method utilizes Monte Carlo simulation. The Sobol' method is employed in this work.

The Sobol' method is well-developed and in wide use in the sensitivity analysis field. Following Homma and Saltelli [5], the main effect sensitivity indices may be estimated via the Sobol' method for a given factor X_i by first estimating the mean f_0 of the function $Y=f(\mathbf{x})$ as

$$\hat{f}_0=\frac{1}{N}\sum_{m=1}^{N}f(\mathbf{x}^m), \tag{3}$$

where $\mathbf{x}^m$ represents the mth realization of the random vector $[X_1,\ldots,X_k]^T$ and N denotes the total number of realizations of the random vector. Then estimating the variance of the function as

$$\hat{V}=\frac{1}{N}\sum_{m=1}^{N}f(\mathbf{x}^m)^2-\hat{f}_0^2, \tag{4}$$

and the single-factor partial variances as

$$\hat{V}_i=\frac{1}{N}\sum_{m=1}^{N}f([x_1^m,\ldots,x_i^m,\ldots,x_k^m]^T)f([\tilde{x}_1^m,\ldots,x_i^m,\ldots,\tilde{x}_k^m]^T)-\hat{f}_0^2, \tag{5}$$

where x_j^m and $\tilde{x}_j^m$ denote different samples of factor X_j and the partial variances can be computed for $i\in\{1,\ldots,k\}$. The main effect sensitivity index for factor X_i can then be computed according to

$$\hat{S}_i=\frac{\hat{V}_i}{\hat{V}}. \tag{6}$$

Here it should be noted that improvements in estimating main effect indices using sampling-based methods have been developed by Saltelli et al. [22] and using regression or emulator-based methods by Lewandowski et al. [23], Oakley and O'Hagan [17], Tarantola et al. [24], Ratto et al. [25], Storlie and Helton [26]. Our focus in this work is on sampling-based approaches, which are commonly used in situations where both main effect and total effect indices are desired [22]. The methodology developed in this paper can readily be applied to the sampling-based techniques of Saltelli et al. [22], however, the development is more accessible in the context of the traditional Sobol' method. Adapting the work of this paper to regression and emulator-based methods is a topic for future work.

As noted previously, these main effect sensitivity indices may be used for factor prioritization by ranking inputs according to their main effect indices, which give the percentage of how much output variability can be expected to be eliminated by fixing a particular input somewhere on its domain. However, this use of global sensitivity analysis for factor prioritization relies on the assumption that a given factor, through future research, can be fixed to some point on its domain. The key contribution of this work is to relax that assumption by considering the amount of variance that can be reduced for a given factor as a random variable rather than assuming the variance to be completely reducible. More precisely, we assume that for a given amount of variance reduction for a factor X_i, there is a corresponding family of allowable distributions, and we calculate an average change in the variance of the model output over this family.

Let X_i^o be the random variable defined by the original distribution for some factor X_i, and X_i' be the random variable defined by a new distribution for factor X_i after some further research has been done. X_i^o and X_i' have corresponding main effect sensitivity indices S_i^o and S_i' respectively. Then we can define the ratio of the variance of factor X_i that is not reduced and the total variance of the original distribution of factor X_i as $\lambda_i=\mathrm{var}(X_i')/\mathrm{var}(X_i^o)$. Assuming further research reduces the variance of factor X_i,[1] it is clear that $\lambda_i\in[0,1]$. Since it cannot be known in advance how much variance reduction for a given factor is possible through further research, λ_i is cast as a uniform random variable Λ_i on [0,1], which corresponds to a maximum entropy distribution given that all we know is the interval in which λ_i will take a value [27].

[1] It is possible that further research could increase the variance of a factor, however, this would suggest that the original characterization of uncertainty was flawed.