Reliability Engineering and System Safety 165 (2017) 222–238

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Different numerical estimators for main effect global sensitivity indices

S. Kucherenko*, S. Song

Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Keywords:
Global sensitivity analysis
Sobol' sensitivity indices
Quasi Monte Carlo
Double loop reordering

ABSTRACT

The variance-based method of global sensitivity indices based on Sobol' sensitivity indices became very popular among practitioners due to its easiness of interpretation. For complex practical problems computation of Sobol' indices generally requires a large number of function evaluations to achieve reasonable convergence. Four different direct formulas for computing Sobol' main effect sensitivity indices are compared on a set of test models for which there are analytical results. Considered test functions represent various types of models that are found in practice. Formulas are based on high-dimensional integrals which are evaluated using Monte Carlo (MC) and Quasi Monte Carlo (QMC) techniques. Direct formulas are also compared with a different approach based on the so-called "double loop reordering" formula. It is found that the "double loop reordering" (DLR) approach shows a superior performance among all methods both for models with independent and dependent variables.

1. Introduction

Global sensitivity analysis (GSA) complements Uncertainty Quantification in that it offers a comprehensive approach to model analysis by quantifying how the uncertainty in model output is apportioned to the uncertainty in model inputs [1–3]. Unlike local sensitivity analysis, GSA estimates the effect of varying a given input (or set of inputs) while all other inputs are varied as well, thus providing a measure of interactions among variables. GSA is used to identify key parameters whose uncertainty most affects the output. This information then can be used to rank variables, fix unessential variables and decrease problem dimensionality. The variance-based method of global sensitivity indices based on Sobol' sensitivity indices became very popular among practitioners due to its efficiency and easiness of interpretation [4,5]. There are two types of Sobol' sensitivity indices: the main effect indices, which estimate the individual contribution of each input parameter or a group of inputs to the output variance, and the total sensitivity indices, which measure the total contribution of a single input factor or a group of inputs including interactions with all other inputs [6].

For models with independent variables there are efficient direct formulas which allow to compute Sobol' indices directly from function values. These formulas are based on high-dimensional integrals which can be evaluated via MC/QMC techniques [1,4,5]. For complex practical problems computation of Sobol' indices generally requires a large number of function evaluations to achieve reasonable convergence. More efficient formulas for evaluation of Sobol' main effect indices using direct integral formulas were suggested by Saltelli [7]. Kucherenko et al. [8,9] developed further Saltelli's approach by suggesting new formula which significantly improves the computational accuracy of Sobol' main effect indices with small values. Sobol' and Myshetskaya [10] and Owen [11] suggested their versions of improved direct formulas. In this work we compare original and existing improved direct formulas. For models with dependent inputs we consider a novel approach for estimation Sobol' indices developed in [12]. We also compare direct formulas using MC estimators based on MC and QMC sampling with the so-called double loop approach on a set of test problems which for which there are analytical results for the values of Sobol' indices. The double loop approach has been discarded in the past as being inefficient in comparison with direct formulas but due to the improvements in the algorithms suggested by Plischke [13] it became an interesting alternative to the direct formulas. Further we call this approach as "double loop reordering" (DLR).

Evaluation of Sobol' main effect indices remains to be an active area of research: we could mention application of RBD [14], an Effective Algorithm for Computing Global Sensitivity Indices (EASI) [15], various metamodelling methods [16–18]. In this paper we focused on the direct formulas for computation of Sobol' indices. However, we also considered one of the metamodelling methods, namely Random Sampling High Dimensional Model Representation (RS-HDMR) from [17].

There were other attempts to compare existing and new direct formulas such as the comparison presented in [19]. Authors suggested the MC estimates for reducing errors associated with spurious correla-

* Corresponding author
E-mail address: s.kucherenko@imperial.ac.uk (S. Kucherenko).

http://dx.doi.org/10.1016/j.ress.2017.04.003
Received 3 June 2016; Received in revised form 14 February 2017; Accepted 11 April 2017
Available online 13 April 2017

tions. Practically these estimates can be more efficient than other known formulas only when MC sampling is used. Quasi-random numbers based on f.e. Sobol' sequences [20] practically have negligible correlations in comparison with MC sampling, hence such estimates will not be beneficial in the context of QMC methods. We also note that there were some other attempts to improve direct formulas for Sobol' main effect indices [21]. A new method for improving the efficiency of the methods for computing Sobol' total sensitivity indices was developed in [22].

This paper is organized as follows. The next section introduces ANOVA decomposition and Sobol' sensitivity indices. In Section 3 we present different estimators of the main effect sensitivity indices. Comparison of the efficiency of different estimators is considered in Section 4. Finally, conclusions are given in Section 5.

2. Sobol' sensitivity indices

Consider the square integrable function $f(x)$ defined in the unit hypercube H^d=[0, 1]^d. The decomposition of $f(\mathbf{x})$

$$f(x) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i=1}^{d}\sum_{j>i}^{d} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots d}(x_1, \ldots, x_d), \quad (1)$$

where

$$f_0 = \int_{H^d} f(x)dx$$

is called ANOVA if conditions

$$\int_{H^d} f_{i_1\ldots i_s} dx_{i_k} = 0$$

are satisfied for all different groups of indices $i_1, \ldots, i_s$ such that $1 \le i_1 < i_2 < \ldots < i_s \le d$. These conditions guarantee that all terms in (1) are mutually orthogonal with respect to integration [4].

The variances of the terms in the ANOVA decomposition add up to the total variance:

$$D = \int_{H^d} f^2(x)dx - f_0^2 = \sum_{s=1}^{d}\sum_{i_1<\cdots<i_s}^{d} D_{i_1\ldots i_s},$$

where components $D_{i_1\ldots i_s} = \int_{H^s} f_{i_1\ldots i_s}^2(x_{i_1}, \ldots, x_{i_s}) dx_{i_1}\ldots dx_{i_s}$ are called partial variances.

Main effect global sensitivity indices are defined as ratios

$$S_{i_1\ldots i_s} = D_{i_1\ldots i_s}/D.$$

Further we will consider sensitivity indices for a single index:

$$S_i = D_i/D.$$

Total partial variances account for the total influence of the factor x_i:

$$D_i^{tot} = \sum_{<i>} D_{i_1\ldots i_s},$$

where the sum $\sum_{<i>}$ is extended over all different groups of indices $i_1, \ldots, i_s$ satisfying condition $1 \le i_1 < i_2 < \ldots < i_s \le d$, $1 \le s \le d$, where one of the indices is equal i [1,4]. The corresponding total sensitivity index is defined as

$$S_i^{tot} = D_i^{tot}/D.$$

Sobol' also introduced sensitivity indices for subsets of variables [3,4]. Consider two complementary subsets of variables y and z:

$$x = (y, z).$$

Let $y = (x_{i_1}, \ldots, x_{i_m})$, $1 \le i_1 < \ldots < i_m \le n$, $K = (i_1, \ldots, i_m)$. Here m is a cardinality of a subset y. The variance corresponding to y is defined as

$$D_y = \sum_{s=1}^{m}\sum_{(i_1<\cdots<i_s)\in K} D_{i_1\ldots i_s}. \quad (2)$$

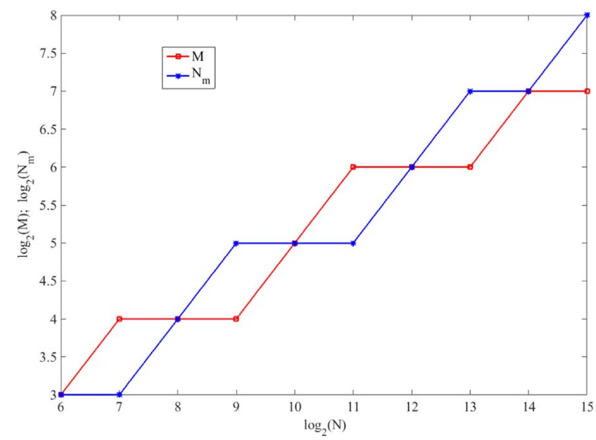

Fig. 1. Dependence of the number of partitions (bins) $log2(M)$ (red line) and sampled points in each partition $log2(N_m)$ (blue line) versus $log2(N)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Number of required function evaluations N_{CPU}.

Method	Sobol'	S-K	Owen	Oracle	DLR	RS-HDMR
Number of function evaluations N_{CPU}	$N(2d+1)$	$N(d+2)$	$N(2d+2)$	$N(d+2)$	N	N

D_y includes all partial variances $D_{i_1}, D_{i_2}, \ldots, D_{i_1\ldots i_s}$ such that their subsets of indices $(i_1, \ldots, i_s) \in K$.

The total variance D_y^{tot} is defined as

$$D_y^{tot} = D - D_z$$

D_y^{tot} consists of all $\sigma_{i_1\ldots i_s}^2$ such that at least one index $i_p \in K$ while the remaining indices can belong to the complementary to K set $\overline{K}$. The corresponding Sobol' sensitivity indices are defined as

$$S_y = D_y/D,$$
$$S_y^{tot} = D_y^{tot}/D.$$

Denote $x_{\sim i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ the vector of all variables but x_i, then $x \equiv (x_i, x_{\sim i})$ and $f(x) \equiv f(x_i, x_{\sim i})$. The first order component in ANOVA decomposition (1) can be found as

$$f_i(x_i) = \int_{H^d} f(x)dx_{\sim i} - f_0.$$

Then

$$D_i = \int_{H^d} [f_i(x_i)]^2 dx_i = \int_{H^d}\left[\int_{H^d} f(x)dx_{\sim i} - f_0\right]^2 dx_i,$$

from which it follows that

$$D_i = \int_{H^d}\left[\int_{H^d} f(x)dx_{\sim i}\right]^2 dx_i - f_0^2. \quad (3)$$

This formula is used to derive a MC estimator known as the brute force estimator or the double loop method.

There is another approach to derive Sobol' sensitivity indices. If we consider x as a random variable uniformly defined in H^d then D_i can be expressed as [1]:

$$D_i = Var_i[E_{\sim i}(f(x_i, x_{\sim i}|x_i))]. \quad (4)$$

This representation can be used to derive an extension of Sobol' sensitivity indices for the case of models with dependent variables [11].