Reliability Engineering and System Safety 134 (2015) 251–259

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Application of the control variate technique to estimation of total sensitivity indices

S. Kucherenko [a,*], B. Delpuech [a], B. Iooss [b], S. Tarantola [c]

[a] Imperial College London, UK
[b] Electricité de France, France
[c] Joint Research Centre of the European Commission, Ispra, Italy

ARTICLE INFO

Available online 17 July 2014

Keywords:
Global sensitivity analysis
Sobol' sensitivity indices
Control variate technique
Random sampling- high dimensional model representation

ABSTRACT

Global sensitivity analysis is widely used in many areas of science, biology, sociology and policy planning. The variance-based methods also known as Sobol' sensitivity indices has become the method of choice among practitioners due to its efficiency and ease of interpretation. For complex practical problems, estimation of Sobol' sensitivity indices generally requires a large number of function evaluations to achieve reasonable convergence. To improve the efficiency of the Monte Carlo estimates for the Sobol' total sensitivity indices we apply the control variate reduction technique and develop a new formula for evaluation of total sensitivity indices. Presented results using well known test functions show the efficiency of the developed technique.

1. Introduction

Global sensitivity analysis (SA) is the study of how the uncertainty in model output is apportioned to the uncertainty in model inputs [1,2]. Over the last decade, global SA has gained acceptance among practitioners in the process of model development, calibration and validation, reliability and robustness analysis, decision-making under uncertainty, quality-assurance, and complexity reduction. There have been many successful improvements in the efficiency of estimating the main effect Sobol' indices ranging from the advanced formulas for small sensitivity indices [3,4] to application of RBD [5] and various metamodelling methods [6–11]. However, there have been no similar advances concerning estimation of the total sensitivity indices and the Sobol–Jansen formula [12,13] remains to be the only formula used in the direct computation of the total sensitivity indices.

To improve the efficiency of the Monte Carlo (MC) estimates for total sensitivity indices we apply the variance reduction technique and develop a new formula for the evaluation of total sensitivity indices. We also present results using well known test functions to show the efficiency of the developed technique.

This paper is organised as follows. The next section introduces control variates reduction technique. ANOVA decomposition and Sobol' sensitivity indices are briefly presented in Section 3. In Section 4 we describe how to apply the control variate technique to improve the efficiency of the Sobol–Jansen formula for evaluation of total sensitivity indices. The use of metamodels is presented in Section 5. This section also provides the details of the random sampling-high dimensional model representation (RS-HDMR) model which is used in this paper. Four different test cases are considered in Section 6. Finally, conclusions are shown in Section 7.

2. Control variate reduction technique

Consider the integral of the function f over the n-dimensional unit hypercube H^n:

$$I[f] = \int f(x)dx.$$

This integral can be seen as an expectation of $f(x)$ with respect to an n-dimensional random variable x that is uniformly distributed:

$$I[f] = E[f(x)] = \int f(x)dx. \tag{1}$$

Further we denote $\mu_f = E[f(x)]$. To calculate this expectation, a sequence of N random or quasi random points $x^{(i)}$ is sampled and then the mathematical expectation (1) is approximated as follows:

$$I_N[f] = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)}) \tag{2}$$

It is important to estimate an integration error $\varepsilon = |I[f] - I_N[f]|$. For the MC method the expectation of ε^2 is equal to $\sigma^2(f)/N$,

* Corresponding author. Tel.: +44 2075946624; fax: +44 2075941129.
E-mail address: s.kucherenko@imperial.ac.uk (S. Kucherenko).

http://dx.doi.org/10.1016/j.ress.2014.07.008
0951-8320/

where $\sigma^2 = Var[f]$ and the root mean square error $\varepsilon_{MC} = \sigma/N^{0.5}$. To reduce ε_{MC} one can either increase the number of sampled points N, or to decrease the variance σ^2. The objective of this paper is to develop a new efficient formula for evaluation of total sensitivity indices based on the reduction of the variance of its MC estimate. There are various variance reduction techniques used to increase the accuracy of the MC estimates [14]. One of them is the control variate method which we further consider.

To improve the MC estimate of μ_f we define a new function $\bar{f}(x)$:

$$\bar{f}(x) = f(x) + C(g(x) - \mu_g), \tag{3}$$

where C is a constant coefficient, $g(x)$ is a function known as a control variate for which an expectation $\mu_g = \int g(x)dx$ is known. It is easy to see that $E[\bar{f}(x)]$ is an unbiased estimator for $\mu_f$for any choice of the coefficient C: $E[f(x)] = E[\bar{f}(x)]$ as $E[C(g(x) - \mu_g)] = 0$.

Consider the variance of the resulting estimator $\bar{f}$:

$$Var[\bar{f}] = Var[f] + C^2 Var[g] + 2C\, Cov[f, g] \tag{4}$$

The main objective is to make sure that $Var[\bar{f}] \le Var[f]$. Minimum of (4)

$$\min_C Var[\bar{f}]$$

is attained at $C = C^*$:

$$C^* = -\frac{Cov[f,g]}{Var[g]}.$$

In this case

$$Var[\bar{f}] = Var[f] - \frac{Cov^2[f,g]}{Var[g]}.$$

The root mean square error ε_{MC} is then equal to $Var[\bar{f}]/N^{0.5}$. The difficulty here is to find a good control variate g to build an unbiased estimator $\bar{f}$ with a reduced variance.

3. Sobol' sensitivity indices

The method of global sensitivity indices is based on the decomposition of a function into summands of increasing dimensionality. Consider an integrable function $f(x)$ defined in the unit hypercube H^n. It can be expanded in the following form:

$$f(x) = f_0 + \sum_{j=1}^{n} f_j(x_j) + \sum_{0 \le i < j \le n} f_{ij}(x_i, x_j) + \ldots + f_{1\ldots n}(x_1, \ldots, x_n). \tag{5}$$

Each of the components $f_{i_1\ldots i_s}(x_{i_1}, \ldots, x_{i_s})$ is a function of a unique subset of variables from x. It can be proven [12,15] that the expansion (6) is unique if

$$\int_0^1 f_{i_1\ldots i_s}(x_{i_1}, \ldots, x_{i_s}) dx_{i_k} = 0, \quad 1 \le k \le s,$$

in which case it is called a decomposition into summands of different dimensions. This decomposition is also known as the ANOVA decomposition.

For square integrable functions, the variances of the terms in the ANOVA decomposition add up to the total variance of the function

$$D = \sum_{s=1}^{n} \sum_{i_1 < \ldots < i_s}^{n} D_{i_1,\ldots,i_s}.$$

Here $D = Var[f]$ is the total variance, $D_{i_1,\ldots,i_s}$ are partial variances defined as

$$D_{i_1,\ldots,i_s} = \int_0^1 f^2_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s}) dx_{i_1}, \ldots, x_{i_s}$$

We note that is customary in the area of global sensitivity indices to use D to denote the variance instead of $Var[f]$ which is more common in statistics.

Sobol' main effect sensitivity indices are defined as the ratios $S_{i_1,\ldots,i_s} = D_{i_1,\ldots,i_s}/D$.

Consider two complementary subsets of variables x_j and $x_{\sim j}$, where $x_{\sim j}$ is $n-1$ dimensional vector. The total variance D_j^T is defined as

$$D_j^T = D - D_{\sim j},$$

where $D_{\sim j}$ is the sum of all the marginal variances not containing j in their subscripts. The corresponding total sensitivity indices [1,12] are equal to

$$S_j^T = D_j^T/D.$$

Obviously, $S_j^T = S - S_{\sim j}$, where $S_{\sim j} = D_{\sim j}/D$. Knowledge of S_j and S_j^T in most cases provides sufficient information to determine the sensitivity of the analysed function to individual input variables.

Sobol' sensitivity indices can be computed using direct formulas [12,13]. Significant progress has been made in improving efficiencies of computing main effect indices [5]–[3]. In the next section we propose a new efficient approach to compute total sensitivity indices.

4. Application of the control variate technique to estimation of total sensitivity indices

We apply the control variate technique idea presented in Section 2 to the evaluation of the total sensitivity indices S_j^T for the case of a single variable x_j. This approach can be easily generalised for the case of a set of variables.

The Sobol–Jansen formula has a form [12,13]:

$$S_j^T = \frac{1}{2D}\int [f(x) - f(x_j', x_{\sim j})]^2 dx\, dx_j', \tag{6}$$

where x_j' is j -th component of the n - dimensional x' vector which is sampled independently from x. In this section we omit for brevity the integral limits.

Application of the control variate technique would require finding a new function (3), such that

$$\bar{F}(x, x_j') = F(x, x_j') + C(G(x, x_j') - \mu_G),$$

where $F(x, x_j') = [f(x) - f(x_j', x_{\sim j})]^2$, $\mu_G = \int G(x, x_j') dx\, dx_j'$. It is not possible to find a control variate $G(x, x_j')$ in a general case. However, a natural choice for the control variate for a function $f(x)$ is to choose the first order terms of ANOVA

$$g(x) = f_0 + \sum_{i=1}^{n} f_i(x_i). \tag{7}$$

Similarly, a good control variate for $f(x_j', x_{\sim j})$ is

$$g(x_j', x_{\sim j}) = f_j(x_j') + \sum_{i=1, i\ne j}^{n} f_i(x_i). \tag{8}$$

Theorem. *Approximating functions $f(x)$ and $f(x_j', x_{\sim j})$ in (6) with (7) and (8) results in the following formula for S_j^T:*

$$S_j^T = \frac{1}{2D}\int [f(x) - f_j(x_j) - [f(x_j', x_{\sim j}) - f_j(x_j')]]^2 dx\, dx_j' + S_j. \tag{9}$$

Proof. Consider function $F_M = f(x) - f_j(x_j) - [f(x_j', x_{\sim j}) - f_j(x_j')]$. Its variance is

$$Var[F_M] = \int [f(x) - f(x_j', x_{\sim j}) - [f_j(x_j) - f_j(x_j')]]^2 dx\, dx_j'$$