Computer Physics Communications 181 (2010) 1212–1217

Contents lists available at ScienceDirect

Computer Physics Communications

www.elsevier.com/locate/cpc

A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices

I.M. Sobol' [a], S. Kucherenko [b,*]

[a] Institute for Mathematical Modelling of the Russian Academy of Sciences, 4 Miusskaya Square, Moscow 125047, Russia
[b] Centre for Process Systems Engineering, Imperial College London, London, SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 7 September 2009
Received in revised form 1 March 2010
Accepted 8 March 2010
Available online 11 March 2010

Keywords:
Global sensitivity analysis
Global sensitivity index
Monte Carlo method
Derivative based global sensitivity measure

ABSTRACT

A new derivative based criterion τ_y for groups of input variables is presented. It is shown that there is a link between global sensitivity indices and the new derivative based measure. It is proved that small values of derivative based measures imply small values of total sensitivity indices. However, for highly nonlinear functions the ranking of important variables using derivative based importance measures can be different from that based on the global sensitivity indices. The computational costs of evaluating global sensitivity indices and derivative based measures, are compared and some important tests are considered.

1. Introduction

The quality of a model depends on a variety of aspects such as accuracy of experimental data, choice of an appropriate model and reliable identification of the unknown model parameters. With regard to these aspects, sensitivity analysis offers a generalized approach for identification of functional dependencies, selection of a model structure from a set of known competing models, effective and reliable identification of important model parameters and input variables and subsequent reduction of model complexity.

Let $x = (x_1, \dots, x_n)$ be a point in the n-dimensional unit hypercube with Lebesgue measure $dx = dx_1 \cdots dx_n$. Consider a model function $f(x_1, \dots, x_n)$ defined in the hypercube H^n. It can be a black box model and not necessarily an analytical expression.

If $f(x) \in L_2$, global sensitivity indices provide adequate estimates for the impact upon $f(x)$ from individual factors x_i or from groups of these factors. Such indices can be efficiently computed by Monte Carlo or quasi-Monte Carlo methods that include values of $f(x)$ at special random or quasi-random points. However, if the number of model evaluations is too high application of global sensitivity indices can become unpractical.

Derivative based importance criteria provide an alternative approach to the same problem: the impact upon $f(x)$ from the factor x_i is estimated by a functional that depends on the derivative $\partial f/\partial x_i$. A link between both approaches was established in [1]. It was proved that derivative based estimates can be successfully used for identifying of non-important factors, but ranking important variables according to their derivative based estimates is not reliable. Numerical experiments show that in certain situations (*e.g.* if the variation of $\partial f/\partial x_i$ is small) derivative based importance estimates can be computed much faster than the corresponding sensitivity indices [2]. And this is the main reason why we are interested in derivative based importance criteria. In other aspects global sensitivity indices are superior to derivative based importance criteria in that they provide more detailed information about models.

In the present paper a new derivative based criterion τ_y is introduced, that is regarded as a possible estimate of the impact upon $f(x)$ from a group of factors $y = (x_{i_1}, \dots, x_{i_s})$. It is proved, that if $f(x)$ is linear with respect to $x_{i_1}, \dots, x_{i_s}$, then the performance of τ_y is equivalent with the performance of the sensitivity index S_y^{tot}. In the case when y consists of one factor x_i, the corresponding criterion τ_i is a slight improvement of the criteria studied in [1].

This paper is organized as follows: The next section contains definitions and main properties of global sensitivity indices. Section 3 introduces the new derivatives based importance criterion τ_y. Section 4 considers the one-factor case criterion τ_i. Sections 5, 6 and 7 contain some simple but significant examples illustrating theoretical results. In Section 8 the case of independent random factors $x_1, \dots, x_n$ is briefly discussed. Finally, conclusions are presented in the last section. Appendix A contains a simple test for arbitrary importance criteria that estimate the impact of groups of variables. It compares the results for global sensitivity indices, the criterion τ_y and the generalization of the Morris measure for groups of variables μ^* proposed in [3].

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

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2010.03.006

Note that global sensitivity indices were introduced in [4]. Their main theoretical properties were developed in [5–10]. Applications of these techniques are presented in *e.g.* [11–13]. Morris elementary effect measures were introduced in [14] and further developed in [3]. Derivative based criteria were introduced in [2]. Their theoretical properties and the link with global sensitivity indices were established in [1]. Numerical examples are given in [2] and [15]. It was shown in [15] that this technique becomes especially efficient if automatic calculation of derivatives is used.

2. Global sensitivity indices

As mentioned before, let $x = (x_1, \ldots, x_n)$ be a point in the n-dimensional unit hypercube with Lebesgue measure $dx = dx_1 \cdots dx_n$. Consider an arbitrary subset of the variables $y = (x_{i_1}, \ldots, x_{i_s})$, $1 \leqslant s < n$, and the set of remaining complementary variables z, so that $x = (y, z)$, $dx = dy\,dz$. All the integrals in the paper are written without integration limits: we assume that each integration variable varies independently from 0 to 1.

2.1. ANOVA-like decomposition

Consider a model function $f(x) \in L_2$. Denote

$$f_0 = \int f(x)\,dx,$$

$$h_1(y) = \int f(x)\,dz - f_0,$$

$$h_2(z) = \int f(x)\,dy - f_0,$$

$$h_{12}(x) = f(x) - \int f(x)\,dz - \int f(x)\,dy + f_0.$$

Then we obtain a decomposition of $f(x)$

$$f(x) = f_0 + h_1(y) + h_2(z) + h_{12}(x). \tag{2.1}$$

One can easily verify that

$$\int h_1(y)\,dy = \int h_2(z)\,dz = \int h_{12}(x)\,dy = \int h_{12}(x)\,dz = 0.$$

These relations imply the orthogonality of the terms in (2.1):

$$\int h_1(y)h_2(z)\,dx = \int h_1(y)h_{12}(x)\,dx = \int h_2(z)h_{12}(x)\,dx = 0.$$

2.2. Variances

The following integrals are called partial variances: $D_y = \int h_1^2(y)\,dy$, $D_z = \int h_2^2(z)\,dz$, $D_{yz} = \int h_{12}^2(x)\,dx$ and D is the total variance: $D = \int f^2(x)\,dx - f_0^2$. Squaring (2.1) and integrating over dx, we obtain $D = D_y + D_z + D_{yz}$. Two total partial variances are defined as $D_y^{tot} = D_y + D_{yz}$, $D_z^{tot} = D_z + D_{yz}$. Note, that if $x_1, \ldots, x_n$ were treated as independent random variables uniformly distributed in the unit interval, then all the terms in (2.1) would be random variables and D, D_y, D_z, D_{yz} their variances.

2.3. Global sensitivity indices

The global sensitivity indices are defined as ratios with common denominator D:

$$S_y = \frac{D_y}{D}, \qquad S_z = \frac{D_z}{D}, \qquad S_{yz} = \frac{D_{yz}}{D},$$

$$S_y^{tot} = \frac{D_y^{tot}}{D}, \qquad S_z^{tot} = \frac{D_z^{tot}}{D}.$$

Obviously: $S_y + S_z + S_{yz} = 1$, $S_y^{tot} = 1 - S_z$, $S_z^{tot} = 1 - S_y$. These indices have the following main properties:

1°. $0 \leqslant S_y \leqslant S_y^{tot} \leqslant 1$.
2°. $S_y = S_y^{tot} = 0$ means that $f(x)$ does not depend on y.
3°. $S_y = S_y^{tot} = 1$ means that $f(x)$ depends only on y.
4°. If the set of variables z is somehow fixed, $z = z_0$, and $f(x)$ is approximated by $f(y, z_0)$, then the approximation error depends strongly on S_z^{tot}.

Here is an exact formulation of property 4° from [8] and [9]. Denote by $\delta(z_0)$ the approximation error in a scaled L_2 metric:

$$\delta(z_0) = \frac{1}{D}\int \left[f(x) - f(y, z_0)\right]^2 dx.$$

Then $\delta(z_0) \geqslant S_z^{tot}$. But if z_0 is random and uniformly distributed, then the expectation is

$$E\delta(z_0) = 2S_z^{tot}.$$

2.4. Integral formulas

The following integral formulas show that global sensitivity indices can be computed without knowing the terms in (2.1):

$$D_y^{tot} = \frac{1}{2}\int \left[f(y', z) - f(x)\right]^2 dx\,dy', \tag{2.2}$$

$$D_y = \int f(x')\left[f(y', z) - f(x)\right] dx\,dx'. \tag{2.3}$$

The integrals on the right-hand side can be computed by crude Monte Carlo or quasi-Monte Carlo methods, the point x' is an independent realization of the point x, so that $x' = (y', z')$ and $dx' = dy'\,dz'$. The total variance D can be computed by traditional statistical methods. If simultaneously m different subset y are considered, then according to [6] for each Monte Carlo trial $m + 2$ model estimates are necessary.

Formula (2.2) according to [5] is a slight generalization of formulas from [16] and [17]. An assertion equivalent to (2.2) was published in [18]. Formula (2.3) was derived in [9]. The variance of the (2.2) estimator was investigated in [5]; the variance of the (2.3) estimator was studied in [19].

3. Importance criterion τ_y

Assume that $f(x)$ is differentiable. Consider the Taylor expansion

$$f(y', z) - f(y, z) = \sum_{p=1}^{s} \frac{\partial f(x)}{\partial x_{i_p}}\left(x'_{i_p} - x_{i_p}\right) + \cdots.$$

The omitted terms depend on second order derivatives. If these derivatives (or more accurately, if certain integrals that include these derivatives) are small, then the last sum will be a fair approximation for the expression on the left-hand side. Substituting this sum into formula (2.2), we construct the following derivative based importance criterion:

$$\hat{\tau}_y = \frac{1}{2}\int \left[\sum_{p=1}^{s} \frac{\partial f(x)}{\partial x_{i_p}}\left(x'_{i_p} - x_{i_p}\right)\right]^2 dx\,dy'.$$

In certain circumstances $\hat{\tau}_y$ will be near to D_y^{tot}. In the formula for $\hat{\tau}_y$, the integration over dy' can be carried out. Indeed,

$$\hat{\tau}_y = \frac{1}{2}\int \sum_{p=1}^{s}\left(\frac{\partial f(x)}{\partial x_{i_p}}\right)^2 \left(x'_{i_p} - x_{i_p}\right)^2 dx\,dy'$$
$$+ \int \sum_{p<q} \frac{\partial f(x)}{\partial x_{i_p}}\frac{\partial f(x)}{\partial x_{i_q}}\left(x'_{i_p} - x_{i_p}\right)\left(x'_{i_q} - x_{i_q}\right) dx\,dy'.$$

The final expression for $\hat{\tau}_y$ is

$$\hat{\tau}_y = \sum_{p=1}^{s} \int \left(\frac{\partial f(x)}{\partial x_{i_p}} \right)^2 \frac{1 - 3x_{i_p} + 3x_{i_p}^2}{6}\, dx + \sum_{p<q} \int \frac{\partial f(x)}{\partial x_{i_p}} \frac{\partial f(x)}{\partial x_{i_q}} \left(x_{i_p} - \frac{1}{2} \right) \left(x_{i_q} - \frac{1}{2} \right) dx.$$

The sum over $p < q$ means that $1 \leqslant p < q \leqslant s$.

τ_y is a further simplification of $\hat{\tau}_y$: only one main term is retained, namely

$$\tau_y = \sum_{p=1}^{s} \int \left(\frac{\partial f(x)}{\partial x_{i_p}} \right)^2 \frac{1 - 3x_{i_p} + 3x_{i_p}^2}{6}\, dx. \tag{3.1}$$

This simplification is justified by the following properties of τ_y.

Theorem 1. *If $f(x)$ is linear with respect to $x_{i_1}, \ldots, x_{i_s}$, then $D_y^{tot} = \tau_y$, or in other words $S_y^{tot} = \frac{\tau_y}{D}$.*

Theorem 2. *A general inequality holds: $D_y^{tot} \leqslant (24/\pi^2)\tau_y$ or in other words $S_y^{tot} \leqslant \frac{24}{\pi^2} \frac{\tau_y}{D}$.*

The first theorem shows that if the model $f(x)$ is near to linear, the performance of τ_y will be near to the performance of global sensitivity indices.

The second theorem shows that small values of τ_y imply small values of S_y^{tot} and this allows identification of a set of unessential factors y (usually defined by a condition of the type $S_y^{tot} < \delta$).

Proof of Theorem 1. Assume that $f(x) = \sum_{p=1}^{s} a_p(z) x_{i_p} + b(z)$. Formula (2.2) then implies

$$D_y^{tot} = \frac{1}{2} \int \left\{ \sum_{p=1}^{s} a_p(z) \left(x'_{i_p} - x_{i_p} \right) \right\}^2 dx\, dy' = \frac{1}{2} \sum_{p=1}^{s} \int a_p^2(z) \left(x'_{i_p} - x_{i_p} \right)^2 dz\, dy' = \frac{1}{12} \int \sum_{p=1}^{s} a_p^2(z)\, dz.$$

On the other hand, from (3.1)

$$\tau_y = \sum_{p=1}^{s} \int a_p^2(z)\, dz \int \frac{1 - 3x_{i_p} + 3x_{i_p}^2}{6}\, dx_{i_p} = \frac{1}{12} \sum_{p=1}^{s} \int a_p^2(z)\, dz. \qquad \square$$

The proof of Theorem 2 will be given at the end of the next section, because this proof needs some results derived in the latter section.

4. Importance criterion τ_i

Consider now the one-dimensional case when the subset y consists of only one variable $y = (x_i)$. From (3.1) we then obtain a criterion

$$\tau_i = \int \left(\frac{\partial f(x)}{\partial x_i} \right)^2 \frac{1 - 3x_i + 3x_i^2}{6}\, dx, \tag{4.1}$$

that is very close to the criterion ν_i, discussed in [1]:

$$\nu_i = \int \left(\frac{\partial f(x)}{\partial x_i} \right)^2 dx. \tag{4.2}$$

In fact, $1 - 3t + 3t^2$ for $0 \leqslant t \leqslant 1$ is bounded: $1/4 \leqslant (1 - 3t + 3t^2) \leqslant 1$. Therefore $\nu_i/24 \leqslant \tau_i \leqslant \nu_i/6$. In [1] a general inequality was proved: $S_i^{tot} \leqslant \frac{1}{\pi^2} \frac{\nu_i}{D}$. From this inequality we immediately obtain a general inequality for τ_i:

$$S_i^{tot} \leqslant \frac{24}{\pi^2} \frac{\tau_i}{D}. \tag{4.3}$$

Thus small values of τ_i imply small values of S_i^{tot}, that are characteristic for non-important variables x_i.

At the same time from Theorem 1 we obtain a corollary: if $f(x)$ depends linearly on x_i, then $S_i^{tot} = \frac{\tau_i}{D}$. Thus τ_i is closer to D_i^{tot} than ν_i.

Note that the constant factor $1/\pi^2$ in the general inequality for ν_i is the best possible. But in the general inequality for τ_i the best possible constant factor is unknown.

4.1. Example

Consider the so-called g-function that is often used in sensitivity analysis for numerical experiments $g(x) = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}$, where the non-negative parameters a_i define the importance of x_i: the larger a_i, the less important the input variable x_i is. For the g-function $D_i^{tot} = d_i \prod_{k \neq i} (1 + d_k)$, where $d_k = \frac{1}{3}(1 + a_k)^{-2}$ (details can be found in [1]). It can be proved that $\tau_i = 4D_i^{tot}$.

Here are two comments in connection with this example.

First, if $\partial^2 f / \partial x_i^2 \equiv 0$, then $f(x)$ is a linear function of x_i, and according to the corollary, $\tau_i = D_i^{tot}$. For the g-function $\partial^2 g / \partial x_i^2 \equiv 0$ for all values of x_i except $x_i = 1/2$, so in this case $\tau_i \neq D_i^{tot}$.

Second, in this example the τ_i are proportional to S_i^{tot} for all x_i; these x_i can be either non-important or very important. This example suggests that the τ_i could be used in the same way as S_i^{tot} for ranking arbitrary variables. The example in the next section shows that this is not always so: ranking important variables according to the values of τ_i and S_i^{tot} can be different.

Proof of Theorem 2. The inequality declared in Theorem 2 in Section 3 follows immediately from the one-dimensional inequality (4.3) and the following two relations:

$$\tau_y = \sum_{p=1}^{s} \tau_{i_p} \tag{4.4}$$

and

$$D_y^{tot} \leqslant \sum_{p=1}^{s} D_{i_p}^{tot}. \tag{4.5}$$

Indeed

$$D_y^{tot} \leqslant \sum_{p=1}^{s} D_{i_p}^{tot} \leqslant \frac{24}{\pi^2} \sum_{p=1}^{s} \tau_{i_p} = \frac{24}{\pi^2} \tau_y.$$

Equality (4.4) follows from the definitions of τ_y and τ_i. Here is a sketch of a proof of the inequality (4.5). We recall the ANOVA decomposition of $f(x)$ into a sum of orthogonal terms of different dimensions and the corresponding decomposition of the total variance D:

$$D = \sum_i D_i + \sum_{i<j} D_{ij} + \cdots + D_{12\cdots n}.$$

The total partial variance D_y^{tot} is a sum of all $D_{k_1\cdots k_m}$ where at least one of the indices $k_j \in (i_1,\dots,i_s)$. The partial variance $D_{i_p}^{tot}$ is the sum of all $D_{k_1\cdots k_m}$ where one of the indices k_j is equal to i_p. Obviously, an arbitrary term $D_{k_1\cdots k_m}$ from D_y^{tot} has an index equal to some i_p, and is included into $D_{i_p}^{tot}$. $\square$

5. Counterexample

Consider a function f which has the following ANOVA decomposition:

$$f = \sum_{i=1}^{4} c_i\left(x_i - \frac{1}{2}\right) + c_{12}\left(x_1 - \frac{1}{2}\right)\left(x_2 - \frac{1}{2}\right)^5,$$

where $c_i = 1$, $1 \leqslant i \leqslant 4$, $c_{12} = 50$. For this function all $S_i = 0.237$, $1 \leqslant i \leqslant 4$, $S_{12} = 0.0523$ and $S_1^{tot} = S_2^{tot} = 0.289$, $S_3^{tot} = S_4^{tot} = 0.237$, so variables 1, 2 have the same importance, and so do variables 3, 4. However, for derivative based importance criteria, variables 1 and 2 have different importance: $\tau_1 = 0.101$, $\tau_2 = 0.409$, while variables 3 and 4 still have equal importance $\tau_3 = \tau_4 = 0.083$. Moreover, $\tau_2 > \tau_1 + \tau_3 + \tau_4$.

One can see that $\tau_i/(\pi^2 D)$ is much higher than S_i^{tot} only for variable 2. It is caused by the strong nonlinearity of the term $f_{1,2}(x_1,x_2)$ with respect to x_2. This example shows that ranking of influential variables based on τ_i may result in false conclusions: in our example x_2 seems more important than all the other variables together.

6. Variances: τ_i versus D_i^{tot}

In this section Monte Carlo algorithms for computing τ_i and D_i^{tot} are compared and the advantage of the algorithm for computing τ_i is shown. Consider a model function $f(x)$ that is linear with respect to x_i:

$$f(x) = a(z)x_i + b(z). \tag{6.1}$$

It follows from Section 4 that for this function $\tau_i = D_i^{tot} = \frac{1}{12}\int a^2(z)\,dz$.

Assume that the integral representation (2.2) is applied for computing D_i^{tot}

$$D_i^{tot} = \frac{1}{2}\int a^2(z)\left(x_i' - x_i\right)^2 dx\,dx_i'.$$

If this integral is estimated by crude Monte Carlo, the estimator's variance is

$$V_1 = \frac{1}{4}\int a^4(z)\left(x_i' - x_i\right)^4 dx\,dx_i' - \left(D_i^{tot}\right)^2.$$

Assume now that formula (4.1) is used for computing τ_i:

$$\tau_i = \int a^2(z)\frac{1 - 3x_i + 3x_i^2}{6}\,dx.$$

The crude Monte Carlo estimator's variance of τ_i is

$$V_2 = \int a^4(z)\left(\frac{1 - 3x_i + 3x_i^2}{6}\right)^2 dx - (\tau_i)^2.$$

Denote by N_1 and N_2 the numbers of Monte Carlo trials (or the sample sizes) required to achieve a prescribed relative standard error:

$$\delta = \frac{1}{D_i^{tot}}\sqrt{\frac{V_1}{N_1}} = \frac{1}{\tau_i}\sqrt{\frac{V_2}{N_2}}.$$

It follows from these relations that $N_1/N_2 = V_1/V_2$.

Assertion 1. *The function (6.1) implies $2 < N_1/N_2 \leqslant 7$, where the maximum 7 corresponds to $a(z) \equiv Const$.*

Proof. Let us introduce an auxiliary parameter $\lambda = (\int a^4(z)\,dz)/(\int a^2(z)\,dz)^2$. Obviously, $\lambda \geqslant 1$. The integrals over dx_i and dx_i' in V_1 and V_2 can be easily computed. Eliminating $\int a^4(z)\,dz$, one obtains: $N_1/N_2 = V_1/V_2 = (\lambda/60 - 1/144)/(\lambda/120 - 1/144)$. One can easily verify that the derivative $d(N_1/N_2)/d\lambda$ is negative in the interval $1 < \lambda < \infty$. Therefore, the maximum value is attained at $\lambda = 1$ and the minimum value is approached as $\lambda \to \infty$. $\square$

It follows from Assertion 1, that the sample size for D_i^{tot} must be several times larger than that for τ_i.

6.1. ν_i versus D_i^{tot}

Assume now that the criterion ν_i is used rather than τ_i with

$$\nu_i = \int a^2(z)\,dz.$$

If this integral is estimated by crude Monte Carlo, the related variance V_3 will be $V_3 = (\lambda - 1)(\int a^2(z)\,dz)^2$. Denote by N_3 the sample size that provides the same relative standard error

$$\delta = \frac{1}{\nu_i}\sqrt{\frac{V_3}{N_3}}.$$

Then we obtain:

$$\frac{N_1}{N_3} = \frac{144V_1}{V_3} = \frac{2.4\lambda - 1}{\lambda - 1}.$$

Assertion 2. *The function (6.1) implies $2.4 < N_1/N_3 \leqslant \infty$ where the maximum ∞ corresponds to $a(z) \equiv Const$.*

Proof. In the interval $1 \leqslant \lambda < \infty$ the ratio N_1/N_3 decreases from ∞ at $\lambda = 1$ to 2.4 as $\lambda \to \infty$. $\square$

It follows from Assertion 2, that if the criterion ν_i is used and λ is close to 1, the required sample size for estimating D_i^{tot} will be by several orders of magnitude larger than the sample size for estimating ν_i. Such an effect has also been numerically observed in [2].

Remark. We have tried to generalize Assertion 1 to nonlinear functions of the type

$$f(x) = a(z)x_i^m + b(z),$$

where $m > 1/2$. However, we have found that at all $m \geqslant 2$, the ratio $N_1/N_2 < 1$. Thus D_i^{tot} is superior to ν_i in terms of the computational effort.

7. Nonlinear example

Consider the function

$$f(x) = \prod_{i=1}^{n} \varphi_i(x_i), \tag{7.1}$$

where $\varphi_i(x_i) = a_i + b_i(x_i - 1/2)$. Obviously, $\int \varphi_i\,dx_i = a_i$, $\int \varphi_i^2\,dx_i = a_i^2 + \varepsilon_i$, where the variance $\varepsilon_i = b_i^2/12$. Denote

$$R_i = \prod_{t\neq i}\left(a_t^2 + \varepsilon_t\right), \qquad R_{ij} = \prod_{t\neq i,j}\left(a_t^2 + \varepsilon_t\right),$$

$$R_{ijk} = \prod_{t\neq i,j,k}\left(a_t^2 + \varepsilon_t\right).$$

7.1. Individual variables

If $y = (x_i)$, then according to Theorem 1, $\tau_y = D_y^{tot}$:

$$D_i^{tot} = \tau_i = \varepsilon_i R_i.$$

Two-dimensional set $y = (x_i, x_j)$.

In this case $\tau_y > D_y^{tot}$:

$$D_y^{tot} = \left(a_i^2 \varepsilon_j + a_j^2 \varepsilon_i + \varepsilon_i \varepsilon_j\right) R_{ij},$$

$$\tau_y = \left(a_i^2 \varepsilon_j + a_j^2 \varepsilon_i + 2\varepsilon_i \varepsilon_j\right) R_{ij}.$$

Three-dimensional set $y = (x_i, x_j, x_k)$.

Here the difference $\tau_y - D_y^{tot}$ increases:

$$D_y^{tot} = \left(a_i^2 a_j^2 \varepsilon_k + a_i^2 a_k^2 \varepsilon_j + a_j^2 a_k^2 \varepsilon_i + a_i^2 \varepsilon_j \varepsilon_k + a_j^2 \varepsilon_i \varepsilon_k + a_k^2 \varepsilon_i \varepsilon_j + \varepsilon_i \varepsilon_j \varepsilon_k\right) R_{ijk};$$

$$\tau_y = \left(a_i^2 a_j^2 \varepsilon_k + a_i^2 a_k^2 \varepsilon_j + a_j^2 a_k^2 \varepsilon_i + 2a_i^2 \varepsilon_j \varepsilon_k + 2a_j^2 \varepsilon_i \varepsilon_k + 2a_k^2 \varepsilon_i \varepsilon_j + 3\varepsilon_i \varepsilon_j \varepsilon_k\right) R_{ijk}.$$

One can see that as the dimension of the set y increases, the difference $\tau_y - D_y^{tot}$ increases too. However, if all $\varepsilon_i \ll a_i^2$, then the main terms in τ_y and D_y^{tot} are the same. One can expect that if in (7.1) the variances $Var(\varphi_i)$ are small, then τ_y will be near to D_y^{tot}.

8. Normally distributed random variables

Consider independent normal random variables $x_1, \ldots, x_n$ with parameters $(a_i; \sigma_i)$. Then instead of the importance criterion τ_i given in (4.1) we now get the expression

$$\tau_i = \frac{1}{2} E\left[\left(\frac{\partial f(x)}{\partial x_i}\right)^2 \left(x_i' - x_i\right)^2\right]. \tag{8.1}$$

The expectation over x_i' can be computed analytically. Then we obtain

$$\tau_i = \frac{1}{2} E\left[\left(\frac{\partial f(x)}{\partial x_i}\right)^2 \frac{(x_i - a_i)^2 + \sigma_i^2}{2}\right]. \tag{8.2}$$

Suppose $f(x)$ is linear with respect to x_i: $f(x) = a(z)x_i + b(z)$. Then $\tau_i = D_i^{tot} = \sigma_i^2 E(a^2(z))$.

From (8.2) we can obtain the following inequality

$$\tau_i \geqslant \frac{\sigma_i^2}{2} E\left[\left(\frac{\partial f(x)}{\partial x_i}\right)^2\right]$$

or

$$\tau_i \geqslant \frac{\sigma_i^2}{2} \nu_i.$$

Using Theorem 4 from [1] stating that $D_i^{tot} \leqslant \sigma_i^2 \nu_i$, we obtain

$$D_i^{tot} \leqslant 2\tau_i$$

or

$$S_i^{tot} \leqslant \frac{2\tau_i}{D}.$$

Using this inequality we can easily prove the following theorem.

Theorem 3. *If $x_1, \ldots, x_n$ are independent normal random variables, then for an arbitrary subset y of these variables*

$$D_y^{tot} \leqslant 2\tau_y,$$

or in other words

$$S_y^{tot} \leqslant \frac{2\tau_y}{D}.$$

9. Conclusions

A new derivative based criterion τ_y for groups of variables is derived. We also introduced a new criterion for a single variable τ_i, which is a modification of the criteria studied in our previous work. A link between global sensitivity indices and new derivative based measures is established. It is shown that for functions linear with respect to a group of variables the performance of τ_y for the group is equivalent to the performance of the sensitivity index S_y^{tot} for the same group. It is proved that small values of derivative based measures imply small values of total sensitivity indices. However, for highly nonlinear functions the ranking of important factors for variance based and derivative based measures can be different.

The computational costs of evaluating global sensitivity indices and derivative based measures for one class of functions were compared. It was shown that for linear and quasi-linear functions derivative based measures require fewer function evaluations, which confirms earlier numerical findings. For highly nonlinear functions, however, evaluations of global sensitivity indices can be cheaper.

10. Practical conclusions

For estimating the impact upon $f(x)$ from a group of variables y, we recommend to use the global sensitivity index S_y^{tot}. If the computation of this index seems too expensive, one can try to apply the derivative based criterion τ_y, taking into account the following facts:

1) If the dependence of $f(x)$ on the variables from y is nearly linear, then the values of τ_y/D are near to the values of S_y^{tot}.
2) In the strongly nonlinear case, small values of τ_y are always significant: they imply small values of S_y^{tot}. At the same time, large values of τ_y are less informative: the ratio τ_y/D can be arbitrarily large, while S_y^{tot} does not exceed 1.

Acknowledgements

The authors would like to thank A. Sobol' for his help in preparation of this manuscript, N. Shah for his support and interest in this work, and anonymous referees for their valuable comments. One of the authors (S.K.) gratefully acknowledges the financial support by the EPSRC grant EP/D506743/1.

Appendix A. A simple test for importance measures

Consider a function $f(x)$ where all the factors $x_1, \ldots, x_n$ are equivalent and do not interact: $f(x) = \sum_{i=1}^{n} u(x_i)$. Here $u(t)$ is an arbitrary function, $0 \leqslant t \leqslant 1$. From general symmetry considerations one can expect that a reasonable importance measure for the set $y = (x_1, \ldots, x_s)$ must be proportional to s.

Global sensitivity indices. Assume that $u(t) \in L_2$. One can easily find out, that $S_y = S_y^{tot} = \frac{s}{n}$.

Importance criterion τ_y. Assume that $u'(t) \in L_2$. From (3.1)

$$\tau_y = s\alpha,$$

where

$$\alpha = \int_0^1 \left[u'(t)\right]^2 \frac{1 - 3t + 3t^2}{6}\, dt.$$

Importance measure μ^*. The very first attempt to define a derivative based importance measure for groups of factors was made in [3].

The importance measure μ^* for the set $y = (x_1, \dots, x_s)$ is defined as $\mu^* = \frac{1}{N}\sum_{k=1}^{N} \eta(x^{(k)})$, where the $x^{(k)}$ are independent trial points, and $\eta(x)$ (so-called "elementary effect") is $\eta(x) = |f(\tilde{x}) - f(x)|/\Delta$, $\tilde{x} = (x_1 \pm \Delta, \dots, x_s \pm \Delta, x_{s+1}, \dots, x_n)$; the signs $+$ and $-$ are random and equiprobable.

Consider the simplest test: $u(t) = t$ (which means that $f = x_1 + \cdots + x_n$). In this case $\eta(x)$ depends neither on x, nor on Δ: it is simply a random variable $\eta = |\pm 1 \pm 1 \cdots \pm 1|$, where the number of units is equal to s. Obviously if $N \to \infty$ the value $\mu^* \xrightarrow{P} E\eta$. Here are the distributions and expectations of η at $s = 1, 2, 3, 4, 5$.

$$s = 1: \quad \eta \equiv 1, \qquad \mathrm{E}\eta = 1,$$

$$s = 2: \quad \eta \equiv \begin{pmatrix} 2 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad \mathrm{E}\eta = 1,$$

$$s = 3: \quad \eta \equiv \begin{pmatrix} 3 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}, \qquad \mathrm{E}\eta = \frac{3}{2},$$

$$s = 4: \quad \eta \equiv \begin{pmatrix} 4 & 2 & 0 \\ \frac{1}{8} & \frac{4}{8} & \frac{3}{8} \end{pmatrix}, \qquad \mathrm{E}\eta = \frac{3}{2},$$

$$s = 5: \quad \eta \equiv \begin{pmatrix} 5 & 3 & 1 \\ \frac{1}{16} & \frac{5}{16} & \frac{10}{16} \end{pmatrix}, \qquad \mathrm{E}\eta = \frac{15}{8}.$$

Clearly, the simplest test for μ^* fails.

References

[1] I.M. Sobol', S. Kucherenko, Derivative based global sensitivity measures and their link with global sensitivity indices, Math. Comput. Simulation 79 (2009) 3009–3017.

[2] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, N. Shah, Monte Carlo evaluation of derivative-based global sensitivity measures, Reliability Engineering and System Safety (2009) 1135–1148.

[3] F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models, Environmental Modelling and Software 22 (2007) 1509–1518.

[4] I.M. Sobol', Sensitivity estimates for nonlinear mathematical models, Mat. Model. 2 (1) (1990) 112–118 (in Russian); English transl.: Math. Model. Comput. Exp. 1 (4) (1993) 407–414.

[5] I.M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulation 55 (1–3) (2001) 271–280.

[6] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Comm. 145 (2002) 280–297.

[7] I.M. Sobol', S.S. Kucherenko, On global sensitivity analysis of quasi-Monte Carlo algorithms, Monte Carlo Methods Appl. 11 (1) (2005) 83–92.

[8] I.M. Sobol', Global sensitivity indices for studying nonlinear mathematical models, Mat. Model. 17 (9) (2005) 43–52;
I.M. Sobol', Global sensitivity indices for studying nonlinear mathematical models, additions, Mat. Model. 19 (11) (2007) 23–24 (in Russian).

[9] I.M. Sobol', S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz, Estimating the approximation error when fixing unessential factors in global sensitivity analysis, Reliability Engineering and System Safety 92 (7) (2007) 957–960.

[10] I.M. Sobol', B.V. Shukhman, On global sensitivity indices: Monte Carlo estimates affected by random errors, Monte Carlo Methods Appl. 13 (1) (2007) 89–97.

[11] A. Saltelli, K. Chan, E.M. Scott (Eds.), Sensitivity Analysis, John Wiley & Sons, 2000.

[12] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice, John Wiley & Sons, 2004.

[13] S. Kucherenko, N. Shah, The importance of being global. Application of global sensitivity analysis in Monte Carlo option pricing, Wilmott (2007) 82–91.

[14] M.D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (1991) 161–174.

[15] A. Kiparissides, S. Kucherenko, A. Mantalaris, E.N. Pistikopoulos, Global sensitivity analysis challenges in biological systems modeling, Industrial Engineering Chemistry Research 48 (2009) 7168.

[16] V. Saltenis, G. Dzemyda, Structure analysis of extremal problems using an approximation of characteristics, Optimal Decision Theory, Vilnius 8 (1982) 124–138 (in Russian).

[17] M.J.W. Jansen, W.A.H. Rossing, R.A. Daamen, Monte Carlo estimation of uncertainty contributions from several independent multivariate sources, in: J. Grasman, G. Van Straten (Eds.), Predictability and Nonlinear Modelling in Natural Sciences and Economics, Kluwer Academic Publishers, Dordrecht, 1994, pp. 334–343.

[18] M.J.W. Jansen, Analysis of variance designs for model output, Comput. Phys. Comm. 117 (1999) 35–43.

[19] I. Sobol', E. Myshetskaya, Monte Carlo estimators for small sensitivity indices, Monte Carlo Methods Appl. 13 (5–6) (2007) 455–465.