Global Sensitivity Indices for Nonlinear Mathematical Models. Review

I.M. Sobol
Institute for Mathematical Modelling
of the Russian Academy of Sciences,
4 Miusskaya Square, Moscow 125047, Russia
E-mail: hq@imamod.ru

S.S. Kucherenko
Imperial College London, SW7 2AZ, UK
E-mail: s.kucherenko@broda.co.uk

Abstract: This is a review of global sensitivity indices that were introduced in I.M. Sobol’ (1990). These indices allow to analyze numerically the structure of a nonlinear function defined analytically or by a “black box”. As an example the Brownian bridge is considered and an example of the application of global sensitivity indices in finance is presented.

Keywords: Monte Carlo, Quasi-Monte Carlo, Global sensitivity analysis, Brownian bridge.

1 What is Global Sensitivity Analysis
Consider the mathematical model described by a function
\[ u = f(x), \]
where the input \( x = (x_1, \ldots, x_n) \) is defined in a certain region \( G \), and the output \( u \) is a real value. Traditional sensitivity analysis that can be called local, is applied to a specified solution, say \( u^* = f(x^*) \). The sensitivity of \( u^* \) with respect to the input can be measured using the derivatives
\[ \frac{\partial f}{\partial x_i} \mid_{x=x^*}. \]

In the global sensitivity approach individual solutions are not considered. The function \( f(x) \) in \( G \) is studied so that the influence of different variables and their subsets, the structure of \( f(x) \) and possible approximations, etc can be analyzed. A. Saltelli, K. Chan and M. Scott (2000).

2 ANOVA- Decomposition
We shall consider square integrable functions \( f(x), x = (x_1, \ldots, x_n) \), defined in the unit hypercube \( 0 \leq x_1 \leq 1, \ldots, 0 \leq x_n \leq 1 \). In the following text integrals written without limits of integration are from 0 to 1 in each variable.

Definition. The representation of \( f(x) \) in a form
\[ f(x) = f_0 + \sum_{i=1}^{n} \sum_{l_1 \cdots l_n} f_{l_1 \cdots l_n} (x_{i_1}, \ldots, x_{i_l}) \]
is called ANOVA-decomposition if
\[ f_0 = \int f(x) \, dx, \]
The variance $A$ global sensitivity index is the ratio of variances $fi = \frac{\sigma_i^2}{\sigma^2}$.

One can easily prove that conditions (2) and (3) define uniquely all the terms in (1). Indeed, integrating (1) over all variables except $x_i$ we obtain

$$\int f(x) \prod_{p \neq i} dx_p = f_0 + f_i(x)$$

Thus all one-dimensional terms $f_i(x)$ are defined. To define the two-dimensional terms $f_{ij}(x_i, x_j)$ we integrate (1) over all variables except $x_i$ and $x_j$:

$$\int f(x) \prod_{p \neq i, j} dx_p = f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j).$$

And so on. The last term $f_{12\ldots n}(x_1, x_2, \ldots, x_n)$ is defined by (1).

An important property of (1) is the orthogonality of its terms:

$$\int f_{1i} f_{1k} dx = 0, \quad \text{if} \ (i_1, \ldots, i_l) \neq (k_1, \ldots, k_l).$$

This is a direct consequence of (3).

### 3 Global Sensitivity Indices

**Definition.** A global sensitivity index is the ratio of variances $S_{i_1 \ldots i_l} = D_{i_1 \ldots i_l}/D$.

It follows from (4) that

$$\sum_{i_1} \sum_{i_1 \leq \ldots \leq i_l} S_{i_1 \ldots i_l} = 1.$$  

Clearly all sensitivity indices are nonnegative, an index $S_{i_1 \ldots i_l}$ is zero if and only if $f_{i_1 \ldots i_l} \equiv 0$.

The following assertion is more or less evident: the function $f(x)$ is a sum of one-dimensional functions if and only if

$$\sum_{i=1}^n S_i = 1.$$  

One-dimensional sensitivity indices $S_i$ were used in some papers for ranking of the input variables $x_i$. However, a more detailed analysis requires the use of total sensitivity indices that will be introduced in the next section.

### 4 Global Sensitivity Indices for Subsets of Variables

Consider an arbitrary subset of variables $x_{i_1}, \ldots, x_{i_k}$, where $1 \leq k_1 < k_2 < \cdots < k_m \leq n$ and $1 \leq m \leq n - 1$. We will denote it by one letter $y = (x_{i_1}, \ldots, x_{i_m})$, and let $z$ be the set of $n - m$ complementary variables; so that $x = (y, z)$. The set of indices $k_1, \ldots, k_m$ will be denoted by $K$.

Two types of sensitivity indices for the set $y$ are introduced:

**Definitions.**

$$S_y = \sum S_{i_1 \ldots i_k},$$

where the sum is extended over all sets $i_1, \ldots, i_k \in K$;

$$S_y^{tot} = \sum S_{i_1 \ldots i_k},$$

where the sum is extended over all sets $i_1, \ldots, i_k$ with at least one index $i_p \in K$; clearly, $0 \leq S_y \leq S_y^{tot} \leq 1$.

The first of the two definitions can be applied for defining $S_y$. Then $S_y^{tot} = 1 - S_y$ and similarly $S_y^{tot} = 1 - S_z$.

An equivalent approach is to introduce a mixed sensitivity index $S_{y,z} = 1 - S_{z,y}$ with at least one index $i_p \in K$; clearly, $0 \leq S_{y,z} \leq S_{y,z}^{tot} \leq 1$.

The most informative are the extreme cases:

A) $S_y = S_y^{tot} = 0$ if and only if the function $f(x)$ does not depend on $y$.

B) $S_y = S_y^{tot} = 1$ if and only if the function $f(x)$ does not depend on $z$ (the function $f(x)$ is assumed to be piecewise continuous).

The set $y$ consists of one variable $y = (x_i)$ then $S_y = S_i$ while $S_y^{tot} = S_i^{tot}$ is the sum of all $S_{i_1 \ldots i_k}$ that contain $i_p = i$.

**Example.**

Let $f = f(x_1, x_2, x_3)$.

If $y = (x_1)$ then $S_y = S_1$, $S_y^{tot} = S_1 + S_12 + S_13 + S_123$.

If $y = (x_2)$ then $z = (x_3)$. Clearly, $S_y = S_1 + S_2 + S_12$. $S_z = S_3$, $S_y^{tot} = S_1 + S_2 + S_12 + S_13 + S_23 + S_123 = 1 - S_3$.

WILMOTT magazine
A major advantage of the theory is the somewhat unexpected fact that it is unnecessary to compute sums in the definitions of $S_y$ and $S_y^l$: both these quantities (or more accurately speaking, the corresponding variances $D_y$ and $D_y^l$) can be computed directly from values of $f(x)$ at specially selected random or quasi-random points.

### 5 Integral Representations for $D_y$ and $D_y^l$

Denote by $D_y$ and $D_y^l$ sums of $D_{i...i}$ that correspond to the sums in the definitions of $S_y$ and $S_y^l$. Then

$$S_y = \frac{D_y}{D}, \quad S_y^l = \frac{D_y^l}{D}.$$  

Let $x$ and $x'$ be independent variables defined in the same hypercube (or consider the product of two hypercubes). Similarly to $x = (y,z)$ we will write $x' = (y',z')$.

**Theorem 1.**

$$D_y = \int f(x)f(y',z') \, dx \, dz - f_0^2.$$  

**Proof.** The integral on the right hand side can be transformed:

$$\int dy \int f(y,z) \, dz \int f(y',z') \, dz' = \int dy \left[ \int f(y,z) \, dz \right]^2.$$  

Substituting (1) into the inner integral and integrating over $dz$ we retain only terms depending on $y$ and $f_0$. They are squared and integrated over $dy$.

Thus, we obtain $D_y + f_0^2$.

**Theorem 2.**

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

**Proof.** The expression on the right hand side is equal to

$$\int f^2(x) \, dx - \int f(x)f(y',z') \, dx \, dy'.$$

According to Theorem 1, this is equal to

$$\int f^2(x) \, dx - (D_y + f_0^2) = D - D_y.$$  

### 6 A Monte Carlo Algorithm

For the $k$-th trial we generate two $m$-dimensional random points $\eta_k$ and $\eta'_k$, and two $(n - m)$-dimensional random points $\zeta_k$ and $\zeta'_k$. Then we compute the function $f(y,z)$ at three points: $f(\eta_k, \zeta_k), f(\eta_k, \zeta'_k)$ and $f(\eta'_k, \zeta_k)$.

Four estimators are computed: $\varphi_k = f(\eta_k, \zeta_k), \varphi_k^2, \psi_k = f(\eta_k, \zeta'_k)$ and $\chi_k = \frac{1}{2} \left[ \varphi_k - f(\eta'_k, \zeta_k) \right]^2$.

After $N$ independent trials at $N \to \infty$

$$\frac{1}{N} \sum_{k=1}^N \varphi_k \rightarrow f_0,$$

$$\frac{1}{N} \sum_{k=1}^N \varphi_k^2 \rightarrow D + f_0^2,$$

$$\frac{1}{N} \sum_{k=1}^N \psi_k \rightarrow D_y + f_0^2,$$

$$\frac{1}{N} \sum_{k=1}^N \chi_k \rightarrow D_y^l.$$  

A quasi-Monte Carlo estimation of $f_0$, $D$, $D_y$ and $D^l$ is also possible. For the trial number $k$ we select one $2n$-dimensional quasi-random point $Q_k = (q_k^1, \ldots, q_k^N)$ and define $\eta_k = (q_k^1, \ldots, q_k^m), \zeta_k = (q_k^{m+1}, \ldots, q_k^n), \eta'_k = (q_{k+m+1}^1, \ldots, q_{k+n}^N), \zeta'_k = (q_{k+m+n+1}^1, \ldots, q_{k+m+n}^n)$. More information on the computation algorithms can be found in I.M. Sobol’ (2001).

### 7 Low Dimensional Approximations of $f(x)$

According to H. Rabitz, O.F. Alis, J. Shorter and K. Shim (1999) very often in mathematical models $f(x)$ low order interactions of input variables have the main impact upon the output. In such cases a low dimensional approximation $f(x) \approx h_L(x), L \ll n$, where

$$h_L(x) = f_0 + \sum_{i=1}^L \sum_{l_i < \cdots < l_i} f_{i_1 \ldots i_L} (x_{i_1} \ldots , x_{i_L})$$


We will use the scaled $I_2$ distance for measuring the error of an approximation $f(x) \approx h(x)$:

$$\delta(f, h) = \frac{1}{D} \int I[f(x) - h(x)]^2 \, dx.$$  

If the crudest approximations $h(x) \equiv \text{const}$ are considered, the best result is obtained at $h(x) \equiv f_0$; then $\delta(f, f_0) = 1$. Hence, good approximations are the ones with $\delta \ll 1$.

**Theorem 3.** If $f(x)$ is approximated by $h_L(x)$, then

$$\delta(f, h_L) = 1 - \sum_{i=1}^L \sum_{l_i < \cdots < l_i} S_{i_1 \ldots i_L}.$$  

**Proof.** The difference $f(x) - h_L(x)$ is squared and integrated:

$$\int [f(x) - h_L(x)]^2 \, dx = \sum_{i=1}^L \sum_{l_i < \cdots < l_i} D_{i_1 \ldots i_L}.$$  

The result is divided by $D$ and the relations (4) and (5) are used.
8 Fixing Unessential Variables

The approximations \( h_i(x) \) of the preceding section were low dimensional but the number \( n \) of variables remained unchanged. Here we consider the case when several of the input variables have little influence on the output. A common practice is to fix somehow these unessential variables. Let \( y \) be the set of important variables and \( z \) the set of complementary ones. The set \( z \) can be called unessential if \( S_i^{\text{tot}} \ll 1 \).

Let \( z_0 \) be an arbitrary value of \( z \) in the \((n - m)\)-dimensional unit hypercube. As an approximation for \( f(x) = f(y, z) \) the function \( h = f(y, z_0) \) can be suggested. The approximation error \( \delta(f, h) \) depends on \( z_0 \) and shall be written as \( \delta(z_0) = \delta(f, h) \). The following theorem shows that \( \delta(z_0) \) is of the order of \( S_i^{\text{tot}} \).

**Theorem 4.** For an arbitrary \( z_0 \)

\[
\delta(z_0) \geq S_i^{\text{tot}},
\]

but if \( z_0 \) is random and uniformly distributed, then for an arbitrary \( \varepsilon > 0 \) with probability exceeding \( 1 - \varepsilon \)

\[
\delta(z_0) \leq \left( 1 + \frac{1}{\varepsilon} \right) S_i^{\text{tot}}.
\]

The proof of Theorem 4 can be found in I.M. Sobol’ (1990). Here we shall only mention a corollary for \( \varepsilon = 1/2 \):

\[
P\{\delta(z_0) < 3S_i^{\text{tot}}\} \geq 0.5.
\]

The very first problem solved with the aid of global sensitivity indices was a technical one. The model depended on 35 variables, and it was defined by a computer code. The designers assumed that 12 of these variables were unessential. They were satisfied when the global sensitivity approach produced the result \( S_i^{\text{tot}} = 0.02 \), here \( z \) is a subset of unessential variables).

9 Improved Computation Schemes

A. Saltelli showed that in problems in which several sensitivity indices are computed simultaneously, the algorithm of Section 6 can be improved. A. Saltelli (2002). The main idea of improvement looks very innocently: both values \( f(x) \) and \( f(x') \) should be used.

**Theorem 1** can be applied to the subset \( z \). Then

\[
D_z = \int f(x')f(y, z') \, dy \, dx' - f_0^2
\]

and

\[
D_y^{\text{tot}} = D - D_z,
\]

therefore both indices \( S_i \) and \( S_i^{\text{tot}} \) can be computed using three values: \( f(x), f(x') \) and \( f(y, z') \). Only one of these values depends on the choice of the set \( y \), while the computational algorithm described in Section 6 included two such values, namely \((y, z')\) and \((y', z')\).

Consider the problem of estimating all one-dimensional indices \( S_i \) and \( S_i^{\text{tot}} \), \( 1 \leq i \leq n \). A Monte Carlo algorithm similar to the one presented in Section 6 which would require \( n + 2 \) model evaluations for each trial: \( f(x), f(x') \) and \( f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \), \( 1 \leq i \leq n \) can be formulated (a direct use of the algorithm from Section 6 would require \( 2n + 1 \) model evaluations.)

Moreover, these \( n + 2 \) model evaluations can be used for computing all two-dimensional indices \( S_i \). Indeed, let \( y_0 \) be the set \((x_i, x_j)\). It follows from Theorem 1 that

\[
D_{y_0} = \int f(x_1, \ldots, x_i, \ldots, x_j, \ldots) \, dx_1 \, dx_j - f_0^2.
\]

Here the omitted variables are the coordinates of \( x' \). The two arguments of \( f \) differ in two positions only, namely \( x_i \) and \( x_j \).

Further, \( D_y = D_{y_0} - D_i - D_j \) and \( S_y = D_y/D \).

For more details see A. Saltelli (2002).

10 Remarks on the Case of Random Input Variables

Assume that \( x_1, \ldots, x_n \) are independent random variables with distribution functions \( F_1(x_1), \ldots, F_n(x_n) \), and \( f(x_1, \ldots, x_n) \) is a random variable with a finite variance

\[
D = \text{Var}(f).
\]

The definition (1) of ANOVA-decomposition remains true but requirements (2) and (3) should be replaced by corresponding expectations:

\[
f_0 = \text{Ef}(x)
\]

and

\[
\int_{-\infty}^{\infty} f_{y_{1,\ldots,i}}(x_i) \, dx_i = 0 \quad \text{for} \quad 1 < p < s.
\]

In this case the variances \( D_{y_{1,\ldots,i}} \) are real variances:

\[
D_{y_{1,\ldots,i}} = \text{Var}(f_{y_{1,\ldots,i}}).
\]

Functional relations that include random variables are true with probability 1.

In F. Campolongo and A. Rossi (2002) it is shown that uncertainty and sensitivity analysis can be valuable tools in financial applications. A delta hedging strategy is analyzed. Considered in the paper financial instrument to be hedged is a caplet, which is an interest rate sensitive derivative. The instrument chosen to hedge the caplet is a forward rate agreement (FRA). The hedging error is defined as the discrepancy between the value of the portfolio at maturity and what it would have been gained investing the initial value of the portfolio at the risk free rate till maturity.

The delta hedging error is considered as a random variable with a certain distribution centered on zero and the target objective function is the 5th percentile of this distribution (VaR). A Monte Carlo experiment is performed in order to obtain the hedging error empirical distribution and to estimate its 5th percentile. Uncertainty analysis is then used to quantify...
the uncertainty in the variable of interest, while sensitivity analysis is used to identify where this uncertainty is coming from, which is what factors are causing the value of the maximum loss to be uncertain.

There are seven factors contributing to the uncertainty in this value. These include: the features of the caplet (resetting time, interest rate agreed at the outset of the contract, tenor), the parameters of the model (the mean reverting parameter and the spot rate volatility), the strategy used to build the hedging portfolio (represented as a trigger factor describing the type of movements in the yield curve with respect to which the portfolio is immunized), and the number of times at which the portfolio is updated.

Results for the first order indices showed that nearly 55% of the output variance was due to interaction effects among factors. For models expectation (with parameters $0;1$). Then adjacent points $\xi(\frac{i}{n}T)$ are estimated by crude Monte Carlo estimators

$$F_n = \frac{1}{N} \sum_{k=1}^{N} F[\xi_{n,k}(t)]$$

that stochastically converge: $I_{n,N} \rightarrow I_n$, here $\xi_{n,k}(t)$ are independent realizations of $\xi_n(t)$.

In Yu.A. Shreider (1996), two algorithms for constructing $\xi_n(t)$ were described. In both algorithms the time interval $0 \leq t \leq T$ is divided into $n$ equal parts and random values of the process $\xi(t)$ at moments $t = \frac{i}{n}T$ are sampled, $1 \leq i \leq n$. Each value $\xi(\frac{i}{n}T)$ requires one random normal variate $\zeta$ (with parameters $0;1$). Then adjacent points $\xi(\frac{i}{n}T), \xi(\frac{i+1}{n}T)$ in the $(t,x)$ plane are connected by straight lines and thus polygonal line $\xi_n(t)$ is constructed.

In the first algorithm which is often called Standard the random values are sampled in the natural order:

$$\xi\left(\frac{1}{n}T\right), \xi\left(\frac{2}{n}T\right), \ldots, \xi\left(T\right).$$

In the second algorithm it is assumed that $n$ is an integer power of 2, and conditional distributions for the middle of a time interval are applied. The order of sampling is

$$\xi(T), \xi\left(\frac{1}{2}T\right), \xi\left(\frac{1}{4}T\right), \xi\left(\frac{3}{4}T\right), \xi\left(\frac{1}{8}T\right), \ldots, \xi\left(\frac{n-1}{n}T\right).$$

The second algorithm became later known as the Brownian bridge P. Jaeckel (2002).

The probability distributions for $\xi_n(t)$ in both algorithms are the same, hence the variances of $F[\xi_n(t)]$ are equal, and the corresponding Monte Carlo estimators are equivalent. However, it was known that in quasi-Monte Carlo implementations the Brownian bridge is superior to the Standard algorithm. References can be found in I.M. Sobol’ and S.S. Kucherenko (2004), where this conclusion was confirmed by sensitivity analysis.

B) Model and its analysis. As a model functional we consider the functional from Yu.A. Shreider 1996:

$$F[x(t)] = \int_0^T x^2(t) \, dt.$$ Assume that $T = 1$, and the diffusion coefficient in the definition of Wiener’s measure is $0.5$. Then $I = \frac{1}{2}$ and the variance $\text{Var}(F[\xi(t)]) = \frac{1}{3}$.

For both algorithms the integral

$$I_n = \text{EF}_n = \frac{1}{N} \sum_{i=1}^{N} a_i$$

can be computed analytically and the result is

$$I_n = \frac{1}{2} \sum_{i=1}^{n} a_i^2 + \sum_{i<j} a_i a_j \zeta_i \zeta_j,$$

where $\zeta_1, \ldots, \zeta_n$ are independent values of $\zeta$. The coefficients $a_i$ and $a_j$ are different for both algorithms despite the fact that the expectation

$$\text{Var}(\text{EF}_n) = 2 \sum_{i=1}^{n} a_i^2 + \sum_{i<j} a_i^2$$

are the same. For example, the coefficients $a_i$ at $n = 4$ are

$$a_1 = \frac{10}{48}, \quad a_2 = \frac{7}{48}, \quad a_3 = \frac{4}{48}, \quad a_4 = \frac{1}{48},$$

for the Standard algorithm and

$$a_1 = \frac{16}{48}, \quad a_2 = \frac{4}{48}, \quad a_3 = \frac{1}{48}, \quad a_4 = \frac{1}{48},$$

for the Brownian bridge.
The ANOVA-decomposition of $F_n$ is

$$F_n = I_n + \sum_i a_i (\zeta_i^2 - 1) + \sum_{i<j} a_{ij} \zeta_i \zeta_j.$$ 

There are one-dimensional and two-dimensional terms only and

$$S_i = \frac{\text{Var} \left[ a_i (\zeta_i^2 - 1) \right]}{\text{Var}(F_n)} = \frac{2a_i^2}{\text{Var}(F_n)}.$$ 

Table 1 from I.M. Sobol’ and S.S. Kucherenko (2004) contains sums of one-dimensional sensitivity indices at different $n$ for both algorithms as well as values of $I_n$ and variances $\text{Var}(F_n)$.

**TABLE 1**

<table>
<thead>
<tr>
<th>$n$</th>
<th>$I_n$</th>
<th>$\text{Var}(F_n)$</th>
<th>$\sum S_i$, Stand</th>
<th>$\sum S_i$, BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.452</td>
<td>0.323</td>
<td>0.4367</td>
<td>0.7207</td>
</tr>
<tr>
<td>8</td>
<td>0.479</td>
<td>0.331</td>
<td>0.2361</td>
<td>0.7214</td>
</tr>
<tr>
<td>16</td>
<td>0.489</td>
<td>0.332</td>
<td>0.1222</td>
<td>0.7215</td>
</tr>
<tr>
<td>32</td>
<td>0.495</td>
<td>0.333</td>
<td>0.0612</td>
<td>0.7215</td>
</tr>
</tbody>
</table>

Clearly the main contribution to $F_n$ in the Brownian bridge comes from one-dimensional terms (approximately 72%), while for the Standard algorithm the role of two-dimensional terms increases with $n$. As a rule, in quasi-Monte Carlo, one-dimensional integrals are evaluated with greater accuracy than integrals of higher dimensions. Therefore, the Brownian bridge is more accurate than the Standard algorithm.

**Final Remark**

In our example, the sensitivity indices for $F_n$ were evaluated analytically. In general, Monte Carlo or quasi-Monte Carlo computations should be used. To avoid a loss of accuracy when $f_0$ is large, use $f(x) - c$ rather than $f(x)$, with an arbitrary $c \approx f_0$.

**REFERENCES**