The principal variance-based sensitivity indices can be estimated by the "Ishigami-Homma-Saltelli" method *without assuming independence* between the input variables

Bernard Krzykacz-Hausmann
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS), GERMANY
E-mail: bernard.krzykacz-hausmann@grs.de

Summary: It is well-known that the original **I**shigami-**H**omma-**S**altelli (I-H-S) sampling method for estimating Sobol's principal variance-based sensitivity indices is valid for independent variables, only. In this paper it is pointed out that nearly the same procedure, just very slightly modified, can be used in case of not independent variables, too.

Terminology:

$Y = h(X_1,X_2,...,X_m)$: model output considered as a scalar function of the inputs

$\mathbf{X}=(X_1,\ldots..,X_m)$ vector of *not necessarily independent* input random variables,

i.e. with a joint multivariate density $f(x_1,\ldots..,x_m) \neq \prod_i f_i(x_i)$

Partition of $\mathbf{X} =(X_1,\ldots..,X_m) = (\mathbf{X}_{(1)} , \mathbf{X}_{(2)})$ into

$\mathbf{X}_{(1)}$ = subset of those variables for which the sensitivity is to be determined

$\mathbf{X}_{(2)}$ = all the other variables (the rest)

$\Rightarrow$ joint density: $f(\mathbf{x}_{(1)},\mathbf{x}_{(2)}) = f(\mathbf{x}_{(2)}|\mathbf{x}_{(1)})\cdot f(\mathbf{x}_{(1)})$

E.g.

$\mathbf{X}_{(1)}=X_i$, $\mathbf{X}_{(2)}=(X_1,\ldots., X_{i-1}, X_{i+1},...,X_m)$ and

$f(x_1,\ldots..,x_m)= f(x_i| x_1,\ldots x_{i-1}, x_{i+1},.., x_m)\cdot f(x_1,\ldots x_{i-1}, x_{i+1},.., x_m)$

Direct definition of the two principal variance-based sensitivity indices for $\mathbf{X}_{(1)}$:

$SM_{(1)}$ = “Main effect sensitivity index” for $\mathbf{X}_{(1)}$ (= correlation ratio^2)

$$= \frac{\text{var}\,E[Y|\mathbf{X}_{(1)}]}{\text{var}Y} = \frac{\text{var}Y - E\,\text{var}[Y|\mathbf{X}_{(1)}]}{\text{var}Y}$$

≈ "relative amount of variance of Y that is expected to be *removed* if the true value(s) of the variable(s) $\mathbf{X}_{(1)}$ would become known".

$ST_{(1)}$ = “Total effect sensitivity index” for $\mathbf{X}_{(1)}$

$$= \frac{E\,\text{var}[Y|\mathbf{X}_{(2)}]}{\text{var}Y} = \frac{\text{var}Y - \text{var}\,E[Y|\mathbf{X}_{(2)}]}{\text{var}Y}$$

≈ "relative amount of variance of Y that is expected to *remain* if the true values of all the other variables $\mathbf{X}_{(2)}$ would become known".

(exchanging $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$ $\Rightarrow$ $ST_{(1)} = 1 - SM_{(2)}$ $\Rightarrow$ only $SM_{(.)}$ will be considered)

Original I-H-S sampling procedure to estimate $SM_{(1)}$ in the independent case

1[st] y-sample	2[nd] y'-sample
$y_1 = h(\mathbf{x}_{(1),1},\mathbf{x}_{(2),1})$	$y'_1 = h(\mathbf{x}_{(1),1},\mathbf{x'}_{(2),1})$
. .	. .
. .	. .
. .	. .
$y_n = h(\mathbf{x}_{(1),n},\mathbf{x}_{(2),n})$	$y'_n = h(\mathbf{x}_{(1),n},\mathbf{x'}_{(2),n})$

where

$\mathbf{x}_{(1),i}$ is sampled from the pdf $f(\mathbf{x}_{(1)})$ of $\mathbf{X}_{(1)}$

$\mathbf{x}_{(2),i}$ is sampled from the pdf $f(\mathbf{x}_{(2)})$ of $\mathbf{X}_{(2)}$

$\mathbf{x'}_{(2),i}$ is sampled from the same pdf $f(\mathbf{x}_{(2)})$ of $\mathbf{X}_{(2)}$ independently of $\mathbf{x}_{(2),i}$

(i.e. $\mathbf{x}_{(2),i}$ and $\mathbf{x'}_{(2),i}$ are sampled *independently* of $\mathbf{X}_{(1)}$ and of each other)

Modified I-H-S sampling procedure to estimate $SM_{(1)}$ in the not independent case

1st y-sample	2nd y'-sample
$y_1 = h(\mathbf{x}_{(1),1},\mathbf{x}_{(2),1})$	$y'_1 = h(\mathbf{x}_{(1),1},\mathbf{x'}_{(2),1})$
. .	. .
. .	. .
. .	. .
$y_n = h(\mathbf{x}_{(1),n},\mathbf{x}_{(2),n})$	$y'_n = h(\mathbf{x}_{(1),n},\mathbf{x'}_{(2),n})$

where

$\mathbf{x}_{(1),i}$ is sampled from the marginal pdf $f(\mathbf{x}_{(1)})$ of $\mathbf{X}_{(1)}$

$\mathbf{x}_{(2),i}$ is sampled from the conditional pdf $f(\mathbf{x}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1),i})$ of $\mathbf{X}_{(2)}$ given "$\mathbf{X}_{(1)}=\mathbf{x}_{(1),i}$"

$\mathbf{x'}_{(2),i}$ is sampled from the same cond. pdf $f(\mathbf{x}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1),i})$ of $\mathbf{X}_{(2)}$ given "$\mathbf{X}_{(1)}=\mathbf{x}_{(1),i}$" independently of $\mathbf{x}_{(2),i}$

(i.e. $\mathbf{x}_{(2),i}$ and $\mathbf{x'}_{(2),i}$ are sampled *conditionally independently* given "$\mathbf{X}_{(1)}= \mathbf{x}_{(1),i}$")

Modified I-H-S estimate of $SM_{(1)}$ from the two samples $\mathbf{y}=(y_1, \ldots y_n)$ and $\mathbf{y'}=(y'_1,\ldots,y'_n)$ in the general not independent case:

From the resulting two samples $\mathbf{y}=(y_1, \ldots ,y_n)$ and $\mathbf{y'}=(y'_1, \ldots,y'_n)$ compute the estimate of $SM_{(1)}$:

$$\hat{SM}_{(1)} = \frac{\frac{1}{n}\sum y_i \cdot y'_i - \bar{y}\cdot\bar{y}'}{\sqrt{\frac{1}{n}\sum (y_i-\bar{y})^2 \cdot \frac{1}{n}\sum (y'_i-\bar{y}')^2}} =$$

$= \text{corr}(\mathbf{y},\mathbf{y'})$

= sample correlation coefficient between **y** and **y'**.

Original I-H-S estimate of $SM_{(1)}$ from the two samples **y**=(y_1,…,y_n) and **y'**=(y'_1,…,y'_n) in the independent case:

$$\hat{SM}_{(1)} = \frac{\hat{D}_1}{\hat{D}} = \ldots\ldots = \frac{\frac{1}{n}\sum y_i \cdot y'_i - \bar{y}^2}{\frac{1}{n}\sum (y_i - \bar{y})^2}$$

Slight modification of the I-H-S estimate:

$$\bar{y}^2 \leftarrow \bar{y}\cdot\bar{y}', \qquad \frac{1}{n}\sum (y_i - \bar{y})^2 \leftarrow \sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2 \cdot \frac{1}{n}\sum (y'_i - \bar{y}')^2}$$

yields

$$\hat{SM}_{(1)} = \frac{\frac{1}{n}\sum y_i \cdot y'_i - \bar{y}\cdot\bar{y}'}{\sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2 \cdot \frac{1}{n}\sum (y'_i - \bar{y}')^2}} = \mathrm{corr}(\mathbf{y},\mathbf{y'}) =$$

= the sample correlation coefficient between **y** and **y'**

= the same as in the general not independent case.

Basic result:

Consider the two outcome variables

$$Y = h(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}),$$

$$Y' = h(\mathbf{X}_{(1)}, \mathbf{X'}_{(2)}),$$

with $\mathbf{X}_{(2)}$ and $\mathbf{X'}_{(2)}$ identically distributed and conditionally independent given $\mathbf{X}_{(1)}$,

i.e. $f(\mathbf{x}_{(2)},\mathbf{x'}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)}) = f(\mathbf{x}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)})\cdot f(\mathbf{x'}_{(2)}|\mathbf{X}_{(1)}=\mathbf{x}_{(1)})$.

Then the "main effect" sensitivity index $SM_{(1)}$ can be represented as the correlation coefficient $\rho(Y,Y')$ between Y and Y', i.e.

$$SM_{(1)} = \frac{\mathrm{var}\, E[Y|\mathbf{X}_{(1)}]}{\mathrm{var} Y} = \frac{E(Y\cdot Y') - EY\cdot EY'}{\sqrt{\mathrm{var}Y\ \mathrm{var}Y'}} = \rho(Y,Y')$$

Proof: (follows easily from the properties of conditional expectation $E[Y|\mathbf{X}]$)

Y and Y' are identically distributed and conditionally independent given $\mathbf{X}_{(1)}$

$\Rightarrow$ $E(Y') = E(Y) = E(E[Y|\mathbf{X}_{(1)}])$, $\text{var}Y = \text{var}Y'$, $E[Y'|\mathbf{X}_{(1)}] = E[Y|\mathbf{X}_{(1)}]$,

$E[Y \cdot Y'|\mathbf{X}_{(1)}] = E[Y|\mathbf{X}_{(1)}] \cdot E[Y'|\mathbf{X}_{(1)}] = E^2[Y|\mathbf{X}_{(1)}]$

$\Rightarrow E(Y \cdot Y') = E(E[Y \cdot Y'|\mathbf{X}_{(1)}]) = E(E[Y|\mathbf{X}_{(1)}] \cdot E[Y'|\mathbf{X}_{(1)}]) = E(E^2[Y|\mathbf{X}_{(1)}])$

$\Rightarrow EY \cdot EY' = E^2(E[Y|\mathbf{X}_{(1)}])$

$$\Rightarrow \rho(Y,Y') = \frac{E(Y \cdot Y') - EY \cdot EY'}{\sqrt{\text{var}Y \ \text{var}Y'}} = \frac{E(E^2[Y|\mathbf{X}_{(1)}]) - E^2(E[Y|\mathbf{X}_{(1)}])}{\sqrt{\text{var}Y \ \text{var}Y'}} = \frac{\text{var}\, E\,[Y|\mathbf{X}_{(1)}]}{\text{var}Y} = SM_{(1)}$$

Final remarks:

1. The same computational effort as in the traditional independent case.

2. Sobol's variance decomposition doesn't hold in the dependent case.
   The sensitivity indices $SM_{(1)}$ and $ST_{(1)}$ may have different properties.
   E.g. "$ST_{(1)} \geq SM_{(1)}$" may no longer hold.

3. Conditional distributions describe dependence in the most general way. However, it may sometimes be (practically, numerically) difficult to determine all the required conditional distributions and to sample from them as required.