Generalized Hoeffding-Sobol Decomposition for Dependent Variables -Application to Sensitivity Analysis [1]

Gaëlle Chastaing[†] Fabrice Gamboa[‡] Clémentine Prieur[†]

1 General context

I am a PhD student at Laboratoire Jean Kuntzmann, and member of the team INRIA-MOISE in Grenoble. My supervisors are Clémentine Prieur, professor at the Université Joseph Fourier at Grenoble, and Fabrice Gamboa, professor at the Université Paul Sabatier at Toulouse. My thesis is in Statistics. My subject "Sensitivity Analysis for dependent input variables" deals with the case of a non linear model built on dependent inputs. In a stochastic approach, the goal is to quantify the variation of the most influent variables, taking dependency into account.

2 Abstract

Global Sensitivity Analysis is a stochatic approach whose objective is to determine a global criterion of variability based on the joint probability distribution of the output and the inputs of a deterministic model. The most usual quantification is the variance-based method. If Y is the output of a deterministic model η, function of a random vector $\mathbf{X} = (X_1, \cdots, X_p) \in \mathbb{R}^p$, $p \geq 1$,

$$Y = \eta(\mathbf{X}),$$

the functional ANOVA decomposition consists in decomposing the output Y of the model into summands of increasing dimension under orthogonality constraints. Following this approach, Sobol [2] introduces the so called Sobol sensitivity indices to estimate the contribution of groups of inputs on a system. However, these indices are constructed on the hypothesis of independent inputs. In case of correlation, the use of Sobol indices might lead to a wrong interpretation. To deal with this issue, we propose to give an exact and unambiguous definition of the functional ANOVA for correlated inputs. Pionnered by Hooker [3] and Stone [4], we show that η can be uniquely decomposed under boundedness assumptions of the joint distribution P_X of the inputs:

$$\begin{aligned} \eta(X_1, \cdots, X_p) &= \sum_i \eta_i(X_i) + \sum_{i,j} \eta_{ij}(X_i, X_j) + \cdots + \eta_{\mathcal{P}_p}(X_1, \cdots, X_p) \\ &= \sum_{u \in S} \eta_u(X_u) \end{aligned} \quad (1)$$

with the hierarchical orthogonality assumption on the summands:

$$\int \eta_u(x_u)\eta_v(x_v) dP_X(\mathbf{x}) = 0 \quad \forall\ v \subset u,\ \forall\ u \quad (2)$$

[†]Université Joseph Fourier, LJK/MOISE BP 53, 38041 Grenoble Cedex, France, **gaelle.chastaing@imag.fr, clementine.prieur@imag.fr**

[‡]Université Paul Sabatier, IMT-EPS, 118, Route de Narbonne, 31062 Toulouse Cedex 9, France, **fabrice.gamboa@univ-tlse.fr**

This hierarchical decomposition leads to the definition of generalized sensitivity indices. Thus, the sensitivity index S_u of order $|u|$ measuring the contribution of X_u into the model is given by:

$$S_u = \frac{V(\eta_u(X_u)) + \sum_{\substack{v \neq \emptyset \\ u \cap v \neq u,v}} Cov(\eta_u(X_u), \eta_v(X_v))}{V(Y)} \tag{3}$$

Our objective is also to study when this decomposition can be applied, i.e. to know when required conditions are checked. We propose to give conditions on Gaussian mixture, and, in case of two inputs, a general form of copula with illustrations belonging to our framework.
Also, we are concerning by the estimation of generalized sensitivity indices. Our approach is to estimate summands of the decomposition first, then to estimate empirically variance and covariance terms implied in (3). This will be exposed in two different ways for small model dimension, and where the inputs of a model are independent pairs of dependent variables. The first relies on projection operators ad resolution of linear system, the second is based on the resolution of a minimization problem under constraints.

References

[1] G. Chastaing, F. Gamboa, C. Prieur : *Generalized Hoeffding-Sobol Decomposition for Dependent Variables -Application to Sensitivity Analysis*, http://hal.archives-ouvertes.fr/hal-00649404/fr/, 2011.

[2] I.M. Sobol : *Sensitivity estimates for nonlinear mathematical models*, Wiley Ed., vol.1, no.4, 1993.

[3] G. Hooker : *Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables*, Journal of Computational and Graphical Statistics, Vol.16, No.3, 2007.

[4] C.J. Stone : *The use of Polynomial Splines and their tensor products in multivariate function estimation*, The Annals of Statistics, Vol.22, No.1, 1994.