A new variance-based sensitivity analysis for models with non-independent variables.

MATIEYENDOU, LAMBONI
[a] *Department DF-ST, University of Guyane, French Guina, France.*
[b] *228-UMR Espace-Dev, France*

Functional ANOVA ([1-3]) and derivative-based FANOVA ([4]) are widely used in statistical modeling, uncertainty quantification and sensitivity analysis (e.g., [4-10]). Such decompositions of functions $f : \mathbb{R}^d \to \mathbb{R}$ have interesting properties when the input variables are independent, such as i) the uniqueness of the decomposition, ii) Sobol' main indices (i.e., S_js) and interaction indices sum up to one, iii) the Shapley effects of inputs (i.e., Sh_js from [11]) satisfy ([12])

$$0 \leq S_j \leq Sh_j \leq S_{T_j} \leq 1 \,,$$

with S_{T_j} the total index of the input $X_j, \; j = 1, \ldots, d$.

For functions with non-independent input variables (i.e., $\mathbf{X} := (X_1, \ldots, X_d)$), dependency models (DMs) allow for extracting the dependency structures of such variables under the statistical and probabilistic framework ([13-14]). Using $(\sim j) := \{1, \ldots, d\} \setminus \{j\}$ and $\mathbf{Z}_{\sim j}$ for a random vector of $d-1$ independent variables, a DM of $\mathbf{X}$ is given by

$$(X_j, \mathbf{X}_{\sim j}) \stackrel{d}{=} (X_j, r_j (X_j, \mathbf{Z}_{\sim j})) \,,$$

where X_j is at the first position, and $\mathbf{Z}_{\sim j}$ represents $\mathbf{X}_{\sim j}$ in that DM. Composing the function of interest by DMs is used for defining the dependent sensitivity indices (DSIs) of X_js and their upper-bounds (i.e., dS_j, dS_{T_j}, UB_j) in [13]. Such indices verify

$$dS_j = \frac{\mathbb{V}\left[\mathbb{E}\left[f(\mathbf{X})|X_j\right]\right]}{\mathbb{V}\left[f(\mathbf{X})\right]}; \qquad 0 < dS_j \leq dS_{T_j} \leq UB_j, \quad \forall\, j \in \{1, \ldots, d\} \,.$$

Despite the main DSIs are always less than the total ones, note that main indices and interactions do not sum up to one in general, leading to some interpretability issues. It is also the case in [15].

In this abstract, we propose new DSIs that improve the above approach by accounting for the effects of innovation variables Z_js, which represent X_js in some DMs. Basically, our approach consists in collecting necessary and sufficient equivalent representations of $f(\mathbf{X})$ in one multivariate outputs, and then applying the first-type generalized sensitivity indices ([7,8,10]) to assess the effects of X_js. The new main, interaction and total DSIs (i.e., DS_j, DS_u, DS_{T_j}) share the following properties:

$$0 \leq DS_j \leq DS_{T_j} \leq 1; \qquad \sum_{\substack{u \subseteq \{1,\ldots,d\} \\ |u|>0}} DS_u = 1 \,.$$

Note that dS_js are DS_js when neglecting the effects of innovation variables. Also, when all the inputs are independent, we have $DS_j = dS_j = S_j$ and $DS_{T_j} = dS_{T_j} = S_{T_j}$. Our new approach can cope with every model and every distribution of the inputs. For linear models evaluated at the Gaussian random vector, Theorem 1 gives the (new) main and total DSIs of X_j.

Theorem 1 *Let* $f(\mathbf{X}) = \boldsymbol{\beta}^T \mathbf{X}$ *with* $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$*. If* $\Sigma \in \mathbb{R}^{d\times d}$ *has full rank, then*

$$DS_j = DS_{T_j} = \frac{1}{d\,\mathbb{V}[Y]} \sum_{u \subseteq (\sim j)} \binom{d-1}{|u|}^{-1} \frac{\mathbb{C}_{ov}\left[X_j, \mathbf{X}_{\sim u}^T \boldsymbol{\beta}_{\sim u} \,|\, \mathbf{X}_u\right]^2}{\mathbb{V}\left[X_j \,|\, \mathbf{X}_u\right]} .$$

Proof. *Given a matrix* $\mathcal{L}_{\sim\{u,j\},\sim\{u,j\}}$*, such results rely on a DM of* $(\mathbf{X}_u, X_j, \mathbf{X}_{\sim\{u,j\}})$*, that is,*

$$\begin{array}{c} X_j \stackrel{d}{=} \Sigma_{j,u}\left(\Sigma_{u,u}\right)^{-1}\mathbf{X}_u + \Sigma_{j|u}^{1/2} Z_j \\ \mathbf{X}_{\sim\{u,j\}} \stackrel{d}{=} \Sigma_{\sim\{u,j\},u}\left(\Sigma_{u,u}\right)^{-1}\mathbf{X}_u + \Sigma_{\sim\{u,j\},j|u}\Sigma_{j|u}^{-1/2} Z_j + \mathcal{L}_{\sim\{u,j\},\sim\{u,j\}}\mathbf{Z}_{\sim\{u,j\}} \end{array} . \qquad \square$$

In view of Theorem 1, the proposed DSIs are exactly the Shapley effects of Gaussian inputs using linear models (see [11]).

References:

[1] W. Hoeffding, A class of statistics with asymptotically normal distribution, Annals of Mathematical Statistics 19, 293–325, 1948.

[2] B. Efron, C. Stein, The jacknife estimate of variance, The Annals of Statistics 9 586–596, 1981.

[3] F. Kuo, I. Sloan, G. Wasilkowski, H. Woźniakowski, On decompositions of multivariate functions, Mathematics of computation 79 (270), 953–966, 2010.

[4] M. Lamboni, Weak derivative-based expansion of functions: ANOVA and some inequalities, Mathematics and Computers in Simulation 194, 691–718, 2022.

[5] I. M. Sobol, Sensitivity analysis for non-linear mathematical models, Mathematical Modelling and Computational Experiments 1, 407–414, 1993.

[6] A. Saltelli, K. Chan, E. Scott, Sensitivity Analysis, John Wiley and Sons, 2000.

[7] M. Lamboni, H. Monod, D. Makowski, Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models, Reliability Engineering and System Safety 96, 450–459, 2011.

[8] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, Sensitivity indices for multivariate outputs, Comptes Rendus Mathematique 351 (7), 307–310, 2013.

[9] R. Ghanem, D. Higdon, H. Owhadi, Handbook of Uncertainty Quantification, Springer, 2017.

[10] M. Lamboni, Multivariate sensitivity analysis: Minimum variance unbiased estimators of the first-order and total-effect covariance matrices, Reliability Engineering & System Safety 187, 67 – 92, 2019.

[11] A. B. Owen, C. Prieur, On shapley value for measuring importance of dependent inputs, SIAM/ASA Journal on Uncertainty Quantification 5 (1), 986–1002, 2017.

[12] B. Iooss, C. Prieur, Shapley effects for sensitivity analysis with correlated inputs: Comparisons with Sobol' indices, numerical estimation and applications, IJUQ 9 (5), 493–514, 2019.

[13] M. Lamboni, S. Kucherenko, Multivariate sensitivity analysis and derivative-based global sensitivity measures with dependent variables, Reliability Engineering & System Safety 212, 107519, 2021.

[14] M. Lamboni, Efficient dependency models: Simulating dependent random variables, MCS, 2022.

[15] T. Mara, S. Tarantola, P. Annoni, Non-parametric methods for global sensitivity analysis of model output with dependent inputs, EMS 72, 173-183, 2015.

[Matieyendou Lamboni; DFR-ST, UG, matieyendou.lamboni[at]gmail.com/ univ-guyane.fr]