RESEARCH SUBJECT for CEMRACS 2017

Bertrand Iooss, Roman Sueur - EDF R&D Chatou
Véronique Maume-Deschamps – Université Lyon I
Clémentine Prieur – Université Grenoble

Development and estimation of Shapley effects for sensitivity analysis of model outputs

The global sensitivity analysis of a numerical model consists in quantifying the contributions of each of its input parameters in the variability of its output [4]. Based on the functional variance analysis, the popular Sobol' indices [4,10,12] present a faulty interpretation in the presence of correlations between inputs (see the difficulties which arise in [2,6,13]). Recently introduced in this context, the Shapley effects [8], which consist of allocating a part of the variance of the output at each input, make it possible to solve this problem [5,9,13].

The objective of this project is to go beyond on the study of estimation procedures and interpretation of Shapley effects for global sensitivity analysis of computer models with dependent variables. More precisely, the problems to solve are the following:

- Numerical tests: a software implementation (in the package 'sensitivity' of R) has recently been proposed to estimate Shapley effects on numerical models. It would then be useful to start the work by some tests
    - on elementary analytical functions with simple dependence structure between its inputs, in order to give sense to the interpretation of the Shapley effects in terms of variance-based sensitivity indices (comparisons with analytical results obtained in [9] would also be useful),
    - and on an industrial test case that will be proposed by EDF, in order to demonstrate the relevance and usefulness of such sensitivity measures.

- Estimation algorithms:
    - Improvement of the numerical estimation algorithms presented in [13] and estimation of confidence intervals on the Shapley effects (by central limit theorem and/or bootstrap procedures);
    - Development of new estimation algorithms based on a bibliographic search on the Shapley values (e.g. coming from game theory, [1]);
    - Study of the potential links with some random forest-based importance measures which present an equivalence with total Sobol' indices [3];
    - In the case of a computationally expensive model, the direct estimation algorithms of [13] would not be feasible. [5] shows that a metamodel-based algorithm allows to estimate the indices, but without error control. Using the Gaussian process metamodel [11], it would be possible to estimate the Shapley effects by integrating the metamodel uncertainty (as made in [7] for Sobol' indices).

- Development of another version of Shapley effects:
    - For interaction effects;
    - Based on quantile (or superquantile) of the model ouput instead of its variance. This subject seems really open and somewhat risky. Some helps will certainly be necessary from researchers on quantitative finance and risk models.

Références

1. A. Araar and J-Y. Duclos. An algorithm for computing the Shapley Value, 2009. http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf.
2. G. Chastaing, F. Gamboa, and C. Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables - Application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.
3. B. Gregorutti. *Forêts aléatoires et sélection de variables : analyse des données des enregistreurs de vol pour la sécurité aérienne*. Thèse de l'Université Paris VI, France, 2015.
4. B. Iooss and P. Lemaître, A review on global sensitivity analysis methods, In: C. Meloni, G. Dellino (Eds.), Uncertainty management in Simulation- Optimization of Complex Systems: Algorithms and Applications, Springer, 2015, pp. 101{122.
5. B. Iooss et C. Prieur. Analyse de sensibilité avec entrées dépendantes : estimation par échantillonnage et par métamodèles des indices de Shapley, Soumis aux *49èmes Journées de la SFdS*, Avignon, France, Juin 2017.
6. S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183:937-946, 2012.
7. L. Le Gratiet, C. Cannamela, and B. Iooss. A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes. *SIAM/ASA Journal of Uncertainty Quantification*, 2:336–363, 2014.
8. A.B. Owen. Sobol' indices and Shapley value. *SIAM/ASA Journal of Uncertainty Quantification*, 2:245-251, 2014.
9. A.B. Owen and C. Prieur. On Shapley value for measuring importance of dependent inputs, *SIAM/ASA Journal of Uncertainty Quantification*, in press, https://hal.archives-ouvertes.fr/hal-01379188
10. C. Prieur, S. Tarantola, Variance-based sensitivity analysis: Theory and estimation algorithms, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Springer Handbook on Uncertainty Quanti_cation, Springer, 2017.
11. C. Rasmussen, C. Williams, Gaussian processes for machine learning, MIT Press, 2006.
12. I. Sobol, Sensitivity estimates for non linear mathematical models, Mathematical Modelling and Computational Experiments 1 (1993) 407-414.
13. E. Song, B.L. Nelson, and J. Staum. Shapley effects for global sensitivity analysis:Theory and computation. S*IAM/ASA Journal of Uncertainty Quantification*, 4:1060-1083, 2016.