An intuitive variance based variable screening method

R. Sala
Universita degli Studi di Firenze

Affiliation: *Dipartimento di Ingegneria Industriale*, Universita degli Studi di Firenze, Via di Santa Marta 3, 50139 Firenze, Italy

Email: `ramses.sala@unifi.it` – **URL**: `http://www.dief.unifi.it`

Master: Technische Universitaet Muenchen

Ph.D. (2012-2015): Universita degli Studi di Firenze

Supervisor(s): Prof. M. Pierini (Universita degli Studi di Firenze) and Prof. N. Baldanzini (Universita degli Studi di Firenze)

Abstract:

Variance based sensitivity analysis methods are a valuable tool for modelers in engineering and other disciplines to quantitatively determine the strength of the relation between input factors and model outputs. There is a great variety of variance based sensitivity methods [Slt10]. For some complex simulation based models (such as for example full vehicle crashworthiness simulation) it can be infeasible to apply established quantitative variance based sensitivity because the models are to computationally expensive to perform enough function evaluations that would meet the sample requirements. For such expensive models qualitative screening methods can be applied instead. In [Cmp11] a unified approach was presented for variable screening using a radial design sampling strategy to estimate the elementary effects, and for an increasing number of samples the total sensitivity indices based could be estimated.

In this communication a simple and intuitive variance based variable screening method is presented and applied to test functions and multidisciplinary automotive simulation models. Although provided enough samples, the approach could approximate the first order sensitivity indices with arbitrary close accuracy; the main reason to bring this approach under attention is for variable screening purposes, when the number of samples is too small to apply established quantitative first order index estimation methods such as EFAST [Stl99], RBD [Trn06] and EASI [Pls10]. Similarly to the previously referenced unified method [Cmp11], also this method enables a seamless extension of the qualitative variable screening to a quantitative sensitivity analysis, all by itself or more efficient when combined with for example the EASI method.

The general idea of the approach is to estimate for each variable the first order contribution to the output variance, by approximating conditioned means and their variances, over discrete intervals of finite size in scatterplot projections. For a high number of intervals the method converges to the definition of the first order effect sensitivity index or main effect index given in [Sbl93]. For a low number of discrete intervals the resulting indices can be too distantly related to the sensitivity indices to label them sensitivity index estimates. The theoretical values towards which the resulting indices, for fixed intervals converge when the number of sample point increases are however extensive properties of the model with respect to the fixed set of intervals chosen. The approach allows selecting the degree of spatial discretization to which the first order sensitivities are approximated. For the remaining of this document the converged values of those (degenerate) indices will be called Discrete Interval based Sensitivity indices (DIS) indices.

For a realistic case with a limited fixed number of samples available, that is insufficient to estimate first order indices without a large statistical error, a compromise has to be made. Smaller intervals, lead to theoretical DIS indices that are closer to the sensitivity indices, but since the number of samples per interval is low the accuracy of the estimation will be low. Whereas few but large intervals contain more samples per interval and the resulting DIS index estimations are more accurate estimates for those DIS indices that correspond to the larger intervals. The tradeoff between statistical accuracy and spatial discretization resolution can be selected.

The approach is extendable for higher order interaction effects, by creating subsequent projections and divisions in discrete intervals. However, inevitably smaller discrete intervals and subintervals require a larger sample density to obtain relevant results, and therefore only interactions of second are considered. The method is tested to work for data obtained with sampling strategies that result in evenly distributed sample spacing, with a low correlation such as obtained by pseudo-random and quasi-random sampling methods, and combinations thereof.

An intuitive overview of the method will be provided, followed by case studies using a high dimensional instance of the Sobol-G function and vehicle simulation models. Convergence comparisons with quantitative estimation methods for first order sensitivity indices will be provided, and the screening effectiveness will be compared with the modified elementary effect method and regression methods. Besides presenting the practical value of coarse discrete interval based variance contribution screening, the intuitive principle of the method could also lower the threshold to start using variance based sensitivity analysis methods, for practitioners that are less experienced in the field of sensitivity analysis.

[Cmp07] Campolongo, Francesca, Jessica Cariboni, and Andrea Saltelli. ”An effective screening design for sensitivity analysis of large models.” Environmental modelling and software 22.10 (2007): 1509-1518.

[Cmp11] Campolongo, Francesca, Andrea Saltelli, and Jessica Cariboni. ”From screening to quantitative sensitivity analysis. A unified approach.” Computer Physics Communications 182.4 (2011): 978-988.

[Sbl93] Sobol, Ilya M. ”Sensitivity estimates for nonlinear mathematical models.” Mathematical Modelling and Computational Experiments 1.4 (1993): 407-414.

[Slt10] Saltelli, Andrea, and Paola Annoni. ”How to avoid a perfunctory sensitivity analysis.” Environmental Modelling & Software 25.12 (2010): 1508-1517.

[Stl99] Saltelli, Andrea, Stefano Tarantola, and KP-S. Chan. ”A quantitative model-independent method for global sensitivity analysis of model output.” Technometrics 41.1 (1999): 39-56.

[Trn06] Tarantola S., Gatelli, D. and T. Mara (2006) Random Balance Designs for the Estimation of First Order Global Sensitivity Indices, Reliability Engineering & System Safety, 91:6, 717-727 Morris, Max D. ”Factorial sampling plans for preliminary computational experiments.” Technometrics 33.2 (1991): 161-174.

[Pls10] Plischke, Elmar. ”An effective algorithm for computing global sensitivity indices (EASI).” Reliability Engineering & System Safety 95.4 (2010): 354-360.

Short biography – The work is performed in the scope of the task “managing the conflict between lightweight design, NVH and Crashworthiness requirements in vehicle design” within the GRESIMO project that is funded by the EC under grand agreement 290050. The presented variable screening approach resulted as a side product during the activities for this task which the focus is more on multidisciplinary optimization of automotive vehicle structures.