Speaker: Marouane IL IDRISSI

Shapley effects for reliability-oriented sensitivity analysis with correlated inputs

Marouane Il Idrissi
EDF R&D, SINCLAIR AI Lab and Institut de Mathématiques de Toulouse (IMT), France, marouane.il-idrissi@edf.fr

Vincent Chabridon
EDF R&D and SINCLAIR AI Lab, France, vincent.chabridon@edf.fr

Bertrand Iooss
EDF R&D, SINCLAIR AI Lab and IMT, France, bertrand.iooss@edf.fr

Numerical models can be of great help with critical systems for risk and reliability assessment. Tracking and understanding failures of such systems can allow to avoid potentially dramatic consequences. *Reliability-oriented sensitivity analysis* (ROSA) [1] aims at measuring how uncertainties induced by the input probabilistic model influence the variability of (potentially rare) failure events. In particular, a set of methods aim at performing what is called *target sensitivity analysis* (TSA) [2], i.e., at catching the influence of the model's inputs (considering the entire input domain) on the *occurrence* of a failure event.

Formally, for a real-valued numerical model $G(.)$, this implies to consider the ROSA variable of interest $\mathbb{1}_{\{G(X)>t\}}(X)$, where $t \in \mathbb{R}$ is a threshold above which the system is assumed to enter a failure state. The traditional TSA quantity of interest (QoI) is the *failure probability* given by $\mathbb{P}(G(X) > t)$. The main goal of TSA is to quantify the influence of the inputs on the variability of the chosen QoI, after uncertainty propagation.

Traditional global sensitivity analysis (GSA) methods (and subsequently, the common ROSA methods) require an independence assumption on the inputs. An example would be the well-known Sobol' indices, whose interpretation is dramatically altered when inputs are correlated. Recent approaches proposed GSA indices which remain interpretable even when statistical dependency is at stake: the Shapley effects [3]. They leverage the framework of *cooperative game theory*, where the central question is the redistribution of wealth among several players. By analogy with the sensitivity analysis of model output framework, it allowed to define a particular decomposition of the output's variance.

By leveraging this framework, one can obtain TSA indices that can be easily interpretable as shares of meaningful statistics. In the case of variance decomposition, the newly proposed *target Shapley effect* [4] allocated to an input variable $X_j, j = 1, \ldots, d$, can be written as:

$$\text{T-Sh}_j = \frac{1}{d} \sum_{A \subset \{-j\}} \binom{d-1}{|A|}^{-1} \left(\frac{\mathbb{V}\left(\mathbb{E}\left[\mathbb{1}_{\{G(X)>t\}}(X)|X_{A\cup\{j\}}\right]\right) - \mathbb{V}\left(\mathbb{E}\left[\mathbb{1}_{\{G(X)>t\}}(X)|X_A\right]\right)}{\mathbb{V}\left(\mathbb{E}\left[\mathbb{1}_{\{G(X)>t\}}\right]\right)} \right),$$

where $\{-j\} = \{1, \ldots, d\} \setminus \{j\}$, and verifying $\sum_{i=1}^{d} \text{T-Sh}_i = 1$. These indices can be understood as a particular aggregation of "individual", "interaction" and "dependency" effects relative to the Shapley values. They are particularly relevant in the context of TSA where high interactions and induced correlations between inputs tend to blur the global comprehension of the underlying studied failure phenomenon. Fig. 1 provides an illustrative example.

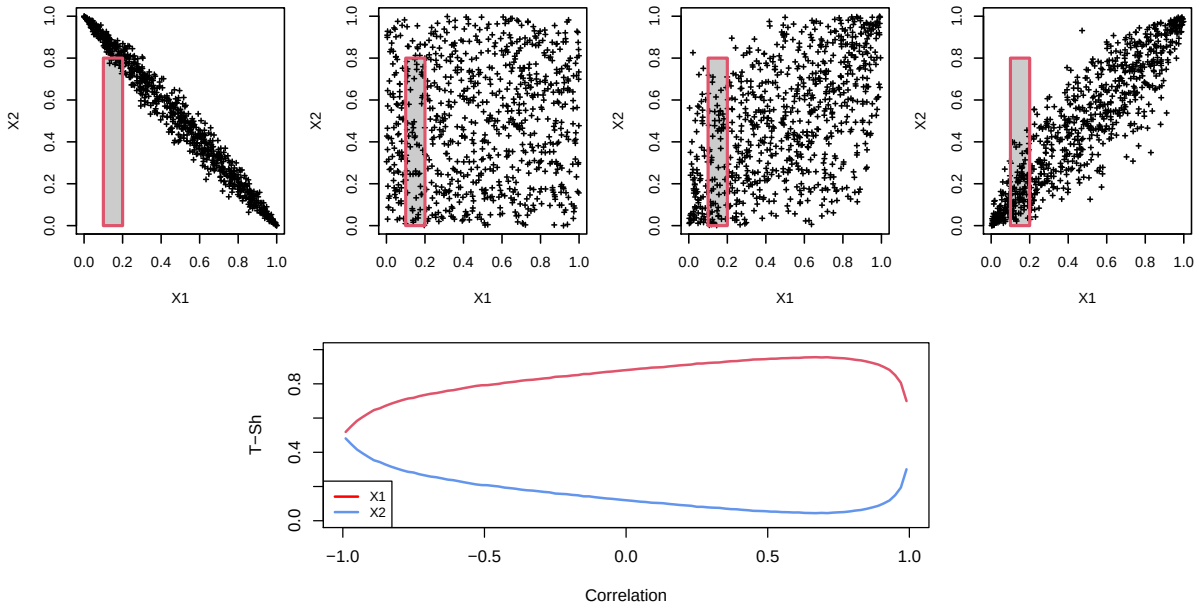

Figure 1: Target Shapley effects for a failure rectangle. The red rectangle represents the failure domain, for different correlation values between the inputs (top row) and each inputs' target Shapley effect with respect to their correlation coefficient (bottom).

Various estimation schemes will be introduced (Monte Carlo sampling as well as a given-data one applicable when a unique data sample is available), with illustrations on analytical results on toy-cases and real industrial applications (e.g., a river flood model and an ultrasonic non-destructive control of a weld).

[1] V. Chabridon. *Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems*. Ph.D. thesis, Université Clermont Auvergne, 2018.

[2] A. Marrel and V. Chabridon Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214: 107711, 2021.

[3] A.B. Owen. Sobol' indices and Shapley value. *SIAM/ASA Journal of Uncertainty Quantification*, 2: 245–251, 2014.

[4] M. Il Idrissi, V. Chabridon, B. Iooss. Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs. *Environmental Modelling & Software*, 143: 105115, 2021.