A new estimator for quantile-oriented sensitivity indices

BROWNE Thomas
Université Paris-Descartes, Paris, France ; EDF Lab Chatou-MRI, Chatou, France

Supervisor(s): Pr. J-C FORT (Université Paris-Descartes), Dr. T. KLEIN (Institut de mathématiques de Toulouse), Dr. L. LE GRATIET (EDF Lab Chatou-MRI) and Dr. B. IOOSS (EDF Lab Chatou-MRI)

Ph.D. expected duration: 2014-2017

Adress: EDF Lab Chatou-MRI, 6 quai Watier, 78401 Chatou, France

Email: thomas.browne@edf.fr

Abstract:

We focus on the sensitivity analysis of casual scalar outputs : we study indices which quantify the influence of a given input over the quantiles of the ouput. In the following we assume that we have Y, the output of a numerical code f, and $(X_1, ..., X_d)$ the scalar random inputs: $Y = f(X_1, ..., X_d)$. The long-term wish for this work is to extend these indices to the case of PoD-curves.

1 Goal oriented-sensitivity analysis (*gosa*)

N. Rachdi introduced the concept of *gosa* in [2]. When sensitivity analysis of Y is required, it states that one must first focus on a property of interest from Y's distribution. The main idea is that, in most cases, only a few probability features of Y are relevant for the study : it could be its mean, $\mathbb{E}[Y]$, any quantile with $\alpha \in]0,1[$, $q^\alpha(Y)$, or a probability of failure, $P_f = \mathbb{P}(Y < 0)$. For instance, if we want to estimate P_f, one could wonder whether all the inputs are influent over it. If one sets the random input X_i to a scalar value x_i, does it change P_f ? We display in Figure 1 a numerical test with 3 inputs, where they are all consecutively set 20 times to a scalar value ($X_i = x_i^j$, $j \in \{1, ..., 20\}$). For each input X_i we compute $\mathbb{P}(Y < 0 \mid X_i = x_i^j)$ and observe its

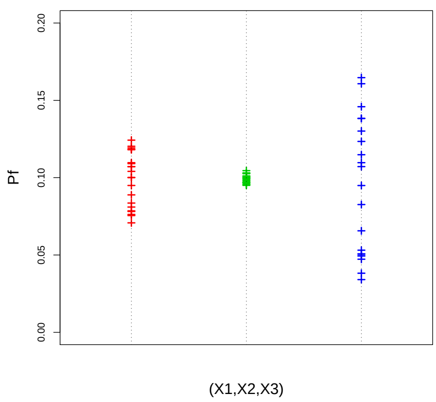

Figure 1: Conditional probabilities of failure P_f: in red, $\mathbb{P}\left(Y \leq 0 \mid X_1 = x_1^j\right)$, in green $\mathbb{P}\left(Y \leq 0 \mid X_2 = x_2^j\right)$ and in blue $\mathbb{P}\left(Y \leq 0 \mid X_3 = x_3^j\right)$, each time for $j = 1, ..., 20$.

variation. In the example one can easily conclude that X_3 has the highest influence over P_f as its variation propagates the most through it. On the other hand, X_2 has no significant influence.

2 Sensitivity analysis indices with respect to a contrast : the quantile case

In [1] the authors introduced the following indice, for $i \in \{1, ..., d\}$ and $\alpha \in]0,1[$:

$$S^i_{c_\alpha}(Y) = \min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y,\theta)\right] - \mathbb{E}_{X_i}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y,\theta) \mid X_i\right]\right],$$

with:

$$\forall y, \theta \in \mathbb{R} \quad c_\alpha(y,\theta) = (y-\theta)(\mathbf{1}_{y\leq\theta} - \alpha),$$

which evaluates the influence of X_i over $q^\alpha(Y)$. Besides, let us recall that $q^\alpha(Y) = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y,\theta)\right]$, therefore:

$$S^i_{c_\alpha}(Y) = \mathbb{E}\left[c_\alpha(Y, q^\alpha(Y))\right] - \mathbb{E}\left[c_\alpha\left(Y, q^\alpha\left(Y \mid X_i\right)\right)\right].$$

Hence one can see that $S^i_{c_\alpha}(Y)$ quantifies the modification of $q^\alpha(Y)$ when one sets X_i to a single value. From a N-sample $\left(Y^1, ..., Y^N\right)$, where $N \in \mathbb{N}$ and for $j \in \{1, ..., N\}$ $Y^j = f\left(X^j_1, ..., X^j_d\right)$, we propose the following kernel-based estimator for $S^i_{c_\alpha}(Y)$:

$$\widehat{S^i_{c_\alpha}(Y)} = \frac{1}{N}\sum_{j=1}^{N} c_\alpha\left(Y^j, \widehat{q^\alpha}(Y)\right) - \frac{1}{hN^2}\sum_{k=1}^{N} \min_{\theta\in\mathbb{R}} \frac{1}{f_i(X^k_i)} \left[\sum_{j=1}^{N} c_\alpha\left(Y^j, \theta\right) K\left(\frac{X^k_i - X^j_i}{h}\right)\right]$$

where K is a kernel, $h > 0$ the bandwidth, f_i the density function of X_i and $\widehat{q^\alpha}(Y)$ the empirical estimator of $q^\alpha(Y)$.

3 Properties of the estimator

Under the conditions $h(N) \underset{N\longrightarrow+\infty}{\longrightarrow} 0$ and $h(N) \times N \underset{N\longrightarrow+\infty}{\longrightarrow} +\infty$, we proved the consistency of the estimator:

$$\lim_{N\longrightarrow+\infty} \widehat{S^i_{c_\alpha}(Y)} = S^i_{c_\alpha}(Y) \quad \text{a.s.}$$

as well as :

$$\sqrt{h(N)\times N}\left(\widehat{S^i_{c_\alpha}(Y)} - S^i_{c_\alpha}(Y) - \mathcal{B}(N)\right) \underset{N\longrightarrow+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \mathcal{V}\right)$$

with $\lim_{N\longrightarrow+\infty} \mathcal{B}(N) = 0$ and $\mathcal{V} > 0$. Numerical tests were performed in order to justify the relevance of this index. Besides, they confirmed the convergence properties.

References

[1] J-C. Fort, T. Klein, and N. Rachdi. New sensitivity analysis subordinated to a contrast. *Communication in Statistics : Theory and Methods*, 2013.

[2] N. Rachdi. *Statistical Learning and Computer Experiments.* PhD thesis, Institut de Mathématiques de Toulouse, France, 2011.

Short biography – I graduated from Université Paris Descartes in 2014 with a masters degree in probability and statistics . I started my PhD at EDF Lab Chatou in November 2014 which is about Probability of Detection curves (PoD curves) for non-destructive tests .