Quantile-oriented sensitivity indices

BROWNE Thomas[1,2], FORT Jean-Claude[2], LE GRATIET Loïc[1]
[1] *EDF Lab Chatou, France*
[2] *Université Paris-Descartes, France*

Goal-oriented sensitivity analysis (*GOSA*, [1])

Let f be a numerical code and Y its one-dimensional output such that $Y = f(X_1, X_2, ..., X_d)$, where $X = (X_1, X_2, ..., X_d)$ are independent random inputs. Regarding a certain strategy for the study, we focus on one precise property of Y's distribution, $\theta(Y)$: it can be $\mathbb{E}[Y], q^\alpha(Y)$, the α-quantile of Y, $\mathbb{P}(Y > t_s)$ with t_s a threshold. If there is a need for sensitivity analysis, *GOSA* states that it may be more relevant to restrict it to $\theta(Y)$. Therefore our wish is to quantify the inputs' influence over $\theta(Y)$. It consists in studying the variability of the conditional parameter $\theta(Y \mid X_i)$. We adopt the following theoritical method: for each input X_i, with $i \in \{1, ..., d\}$, one consecutively sets $X_i = x_i$ for all the possible values of X_i and simulates $f(X_1, ..., x_i, ..., X_d)$ an infinite number of times. Hence one can compute $\theta(Y \mid X_i = x_i)$. We repeat this procedure for all the possible values x_i so that we learn $\theta(Y \mid X_i)$'s distribution. The latter contains the needed information about X_i's influence over $\theta(Y)$.

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

In [2] the authors introduced the following sensitivity index, 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 $q^\alpha(Y) = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y, \theta)\right]$ and $q^\alpha(Y \mid X_i = x_i) = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y, \theta) \mid X_i = x_i\right]$, 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. Besides, we easily prove $\mathbb{E}_{X_i}\left[\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y, \theta) \mid X_i\right]\right] \leq \min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y, \theta)\right]$. Then the authors normalize the index as they divide it by $\min_{\theta\in\mathbb{R}} \mathbb{E}\left[c_\alpha(Y, \theta)\right]$. This now implies: $0 \leq S^i_{c_\alpha}(Y) \leq 1$. In order to justify the meaning of the index, we prove the following property:

$$\begin{array}{lll} S^i_{c_\alpha}(Y) = 0 & \text{if and only if} & q^\alpha\left(Y \mid X_i\right) = q^\alpha(Y) \quad a.s. \\ S^i_{c_\alpha}(Y) = 1 & \text{if and only if} & \forall x_i \quad \mathrm{Var}\left(Y \mid X_i = x_i\right) = 0. \end{array}$$

Estimator and property

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 an estimator for $S^i_{c_\alpha}(Y)$. The first term can be easily estimated by a classical empirical estimation $\min_{\theta\in\mathbb{R}} \frac{1}{n}\sum_{j=1}^n c_\alpha\left(Y^j, \theta\right)$, where $\hat{q}^\alpha(Y) := \arg\min_{\theta\in\mathbb{R}} \frac{1}{n}\sum_{j=1}^n c_\alpha\left(Y^j, \theta\right)$ is the empirical quantile estimator. The second term is much more complicated to estimate as it contains a double expectation (including a conditional expectation) and requires to solve a minimization problem. We base its estimation on the following asymptotic result proved in [3]:

$$\forall x_i \ st \ f_i(x_i) \neq 0, \ \arg\min_\theta \frac{1}{f_i(x_i)} \sum_{j=1}^n c_\alpha\left(Y^j, \theta\right) K_{h(n)}\left(X^j_i - x_i\right) \underset{n\to\infty}{\overset{\mathbb{P}}{\longrightarrow}} \arg\min_\theta \mathbb{E}\left[c_\alpha(Y, \theta) \mid X_i = x_i\right],$$

where K is a positive second-order kernel on a bounded compact, (h_k) the bandwidth sequence and f_i the density function of X_i, with the conditions $h(n) \underset{n\longrightarrow+\infty}{\longrightarrow} 0$ and $h(n) \times n \underset{n\longrightarrow+\infty}{\longrightarrow} +\infty$.

The problem is that we are not interested in the minimizers but in the minimal values for each possible x_i by which we condition, and need to compute their average over the different x_i. At the end, we propose the following kernel-based estimator for $S^i_{c_\alpha}(Y)$:

$$\widehat{S^i_{c_\alpha}(Y)} = \min_{\theta\in\mathbb{R}} \frac{1}{n}\sum_{j=1}^{n} c_\alpha\left(Y^j,\theta\right) - \frac{1}{n}\sum_{k=1}^{n} \min_{\theta\in\mathbb{R}} \frac{1}{k.f_i(X_i^k)}\left[\sum_{j=1}^{k} c_\alpha\left(Y^j,\theta\right)\frac{1}{h_k}K\left(\frac{X_i^k - X_i^j}{h_k}\right)\right].$$

Under the same conditions than above we prove the consistency of the estimator:

$$\widehat{S^i_{c_\alpha}(Y)} \xrightarrow[n\longrightarrow+\infty]{\mathbb{P}} S^i_{c_\alpha}(Y).$$

Applications to defect detection

We study an example in the context of defect examination: we inspect of a structure by sending a wave that reflects on the hypothetical defect. The random reflected signal Z, function of the size of defect a, random environmental properties X and a noise of observation δ, is measured so that: $(Z(a, X, \delta) > t_s)$ implies that the defect is detected. Let us focus on the random defect a_{90}, function of the inputs X, defined as $\mathbb{P}\left(Z(a_{90}, X, \delta) > t_s \mid X\right) = 0.90$, which is the defect that is detected with a probability of 90% under the conditions X. In this example we consider three inputs,

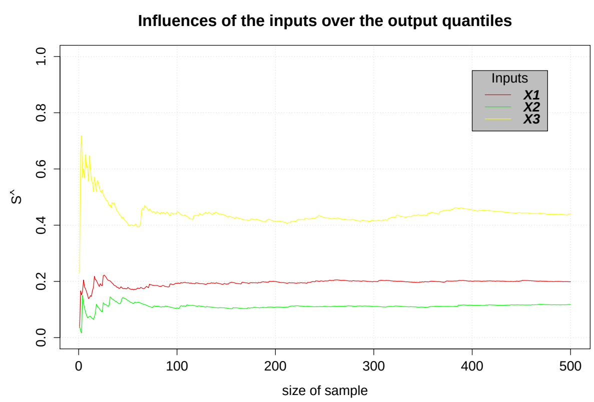

Figure 1: Index estimation for the influence of the three inputs over $q^\alpha(a_{90})$. $\hat{S}^1_{c\alpha}(a_{90})$ is in red, $\hat{S}^2_{c\alpha}(a_{90})$ in green and $\hat{S}^3_{c\alpha}(a_{90})$ in yellow.

$X = (X_1, X_2, X_3)$, and the wish is to estimate $S^1_{c\alpha}(a_{90}), S^2_{c\alpha}(a_{90})$ and $S^3_{c\alpha}(a_{90})$. The different estimators $\hat{S}^1_{c\alpha}(a_{90})$, $\hat{S}^2_{c\alpha}(a_{90})$ and $\hat{S}^3_{c\alpha}(a_{90})$ are computed for a size of sample $n = 2, ..., 150$.

Bibliographie

[1] N. Rachdi (2011), Statistical Learning and Computer Experiments, Thèse de l'Université Paul Sabatier, Toulouse, France.

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

[3] J. Fan, T. Hu and Y. K. Truong (1994), Robust Non-Parametric Function Estimation, *Scandinavian Journal of Statistics, Vol. 21, No. 4, pp. 433-446*

[thomas.ga.browne@gmail.com; EDF R&D, 6 Quai Watier, 78401 Chatou, France]