

**Perturbed-Law based sensitivity Indices for sensitivity analysis in structural reliability**

ROMAN, SUEUR<sup>†</sup>; NICOLAS, BOUSQUET<sup>†</sup>; BERTRAND, IOOSS<sup>†</sup>;  
 JULIEN, BECT<sup>‡</sup>

<sup>†</sup>EDF Lab Chatou, France; <sup>‡</sup>CentraleSupélec, France

In the field of sensitivity analysis of model outputs, the most popular methods are those based on the variance decomposition of the output, such as the Sobol' indices, as they allow defining easy-to-interpret indices measuring the contribution of each input variable to the overall output dispersion. However, these indices are not adapted to the analysis of the impact of inputs on output quantities characterizing extreme events, such as a failure probability, a quantile or a failure domain.

In the following, we denote  $g(\cdot)$  the studied model,  $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{X}$  the random vector of the  $d$  independant input variables and  $Y = g(\mathbf{X})$  the model output.  $\mathbf{X}$  follows a joint probability density function  $f(\mathbf{x})$ . We focus on a typical problem in structural reliability (Morio and Balesdent, 2015), which is the analysis of the failure probability (the failure event arrives with the event  $g(\mathbf{x}) < 0$ ):

$$p = \int_{\mathbb{X}} \mathbf{1}_{\{g(\mathbf{x}) < 0\}} f(\mathbf{x}) d\mathbf{x}.$$

PERTURBED-LAW BASED SENSITIVITY INDICES

Based on the perturbation of each marginal input density, a new type of sensitivity indices has been recently developed (Lemaître et al., 2015). An input  $X_i$ , with marginal density  $f_i$ , is replaced by the perturbed variable  $X_{i\delta}$ .  $X_{i\delta}$  follows the density  $f_{i\delta}$ , based on the initial  $f_i$  perturbed of a  $\delta$  quantity. This allows defining the following perturbed probability  $p_{i\delta}$ :

$$p_{i\delta} = \int_{\mathbb{X}} \mathbf{1}_{\{g(\mathbf{x}) < 0\}} \frac{f_{i\delta}(x_i)}{f_i(x_i)} f(\mathbf{x}) d\mathbf{x}.$$

From this quantity, we define the Perturbed-Law based Indices (PLI) in the following way:

$$S_{i\delta} = \left[ \frac{p_{i\delta}}{p} - 1 \right] \mathbf{1}_{\{p_{i\delta} > p\}} + \left[ \frac{p}{p_{i\delta}} - 1 \right] \mathbf{1}_{\{p_{i\delta} \leq p\}}. \quad (1)$$

The PLI measures have some expected properties, such as being equal to 0 when the failure probability is not changed by the perturbation, or taking a sign that indicates the direction of change of the probability with the  $\delta$  perturbation.  $f_{i\delta}$  is obtained by minimizing the Kullback-Leibler divergence  $KL$  between  $f_{i\delta}$  and  $f_i$  for a given shift  $\delta$  of a statistical characteristic (for example the mean, the variance, a quantile, ...) of the law of  $X_i$ :

$$KL(f_{i,\delta}, f_i) = \int_{-\infty}^{+\infty} f_{i,\delta}(x_i) \log \frac{f_{i,\delta}(x_i)}{f_i(x_i)} dx_i.$$

Easy minimization of  $KL$  provides explicit solutions of  $f_{i\delta}$  for a large range of perturbed parameters of  $f_i$  (e.g. mean, variance, ...) on classical pdf (e.g. Gaussian).

MONTE-CARLO ESTIMATION OF THE PLI MEASURES

An analytical calculation of the probabilities  $p$  and  $p_{i\delta}$  is not possible in practice, and a Monte-Carlo sample  $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N)$  is used. Thus, the estimations are respectively:

$$\hat{p}_N = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{g(\mathbf{x}^n) \leq 0} \quad \text{and} \quad \hat{p}_{i\delta, N} = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{g(\mathbf{x}^n) \leq 0\}} \frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)} f(\mathbf{x}^n).$$

In practice, the estimation of PLI indices might require a very large number of computer experiments if the failure probability is very low (for instance  $10^{-6}$  or less). This could reveal impractical if the code  $g$  is costly.

To improve the estimation of  $p$  and  $p_{i\delta}$ , the importance sampling or subset simulation methods could be used but they often remain too costly in practice (Morio and Balesdent, 2015). We propose here to use a metamodel-based approach which has shown a noticeable efficiency for Sobol' indices (Le Gratiet et al., 2016). In particular, the Gaussian process metamodel allows to control the error on the sensitivity estimates due to the metamodel approximation.

#### BAYESIAN IMPORTANCE SAMPLING

The so-called Bayesian Importance Sampling (BIS) consists in two steps :

- 1) A Gaussian process model  $\tilde{g}$  of the code is built using a budget of  $N_0 < N$  computer experiments. This allows to define a relevant importance density at the following step;
- 2)  $p$  and  $p_{i\delta}$  are estimated with a  $N_1 = N - N_0$  importance sampling scheme (Bect et al., 2015).

The Bayesian optimal importance density  $f_{IS}^*(\mathbf{x})$  is proportional to  $f(\mathbf{x})\sqrt{\mathbb{P}_{N_0}^{\tilde{g}}[g(\mathbf{x}) < 0]}$ . If  $\hat{Z}$  is an estimator of  $Z = \int_{\mathbb{X}} f(\mathbf{x})\sqrt{\mathbb{P}_{N_0}^{\tilde{g}}[g(\mathbf{x}) < 0]}d\mathbf{x}$ , the estimates of  $p$  and  $p_{i\delta}$  are respectively :

$$\hat{p}_{N_0, N_1}^{BIS} = \frac{\hat{Z}}{N_1} \sum_{n=1}^{N_1} \frac{\mathbf{1}_{\{\mathbb{E}_{N_0}^{\tilde{g}}[\tilde{g}(\mathbf{x}^n)] < 0\}}}{\sqrt{\mathbb{P}_{N_0}^{\tilde{g}}[g(\mathbf{x}^n) < 0]}} \quad \text{and} \quad \hat{p}_{i\delta, N_0, N_1}^{BIS} = \frac{\hat{Z}}{N_1} \sum_{n=1}^{N_1} \frac{\mathbf{1}_{\{\mathbb{E}_{N_0}^{\tilde{g}}[\tilde{g}(\mathbf{x}^n)] < 0\}}}{\sqrt{\mathbb{P}_{N_0}^{\tilde{g}}[g(\mathbf{x}^n) < 0]}} \frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)} f(\mathbf{x}^n).$$

#### PERSPECTIVES: PLI FOR SENSITIVITY ANALYSIS OVER A QUANTILE

PLI measures can be applied to an output quantile  $q_\alpha(Y) = \inf\{y \text{ s.t. } F(y) \geq \alpha\}$  with  $F$  the distribution function of  $Y$ . This requires to combine quantile estimation and importance sampling. A naive approach for the estimation of quantiles could consist in replacing  $F$  by its empirical estimator in the latest formula. By re-ordering the  $y^n = g(\mathbf{x}^n)$  we could define  $k_{\alpha, i\delta}$  by

$$k_{\alpha, i\delta} = \min \left\{ k \text{ s.t. } \sum_{n=1}^k \frac{f_{i\delta}(x_i^{(n)})}{f_i(x_i^{(n)})} f(\mathbf{x}^{(n)}) \geq \alpha \right\},$$

where each  $(n)$  denotes the re-ordered index of  $y^n$  in our sample. A straightforward estimator of  $q_\alpha(Y)$  is given by  $\hat{q}_\alpha(Y) = y^{k_\alpha}$ . However, this estimator is unlikely to show good consistency properties, since it is built on a non-continuous quantile function obtained by inverting the empirical cumulative distribution function of  $Y$ . A more promising approach would be to use a quantile-regression framework (Egloff and Leippold, 2010). We notice that  $q_\alpha(Y)$  can be written as  $\operatorname{argmin}_r \mathbb{E}_X[\rho_\alpha(g(X) - r)]$ , where  $\rho_\alpha$  denotes the check-function  $\rho_\alpha(u) = u(\alpha - \mathbf{1}_{\{u < 0\}})$  (see Fort et al., 2013). This suggests the following estimator:

$$\hat{q}_{\alpha, i\delta}(Y) = \operatorname{argmin}_r \sum_{n=1}^N \rho_\alpha(g(\mathbf{x}^n) - r) \frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)} f(\mathbf{x}^n).$$

Replacing  $p$  and  $p_{i\delta}$  in (1) respectively by  $\hat{q}_\alpha$  and  $\hat{q}_{\alpha, i\delta}$  provides the PLI over a quantile.

#### References:

- J. Bect, R. Sueur, A. Géossier, L. Mongellaz, S. Petit and E. Vazquez (2015), *Échantillonnage préférentiel et méta-modèles : méthodes bayésiennes optimale et défensive*, 47èmes Journées de Statistique de la SFdS, Lille, France.
- D. Egloff, M. Leippold (2010), *Quantile estimation with adaptive importance sampling*, Ann. Statist., 38:1244-1278.
- L. Le Gratiet, S. Marelli and B. Sudret (2016), *Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes*, In: *Springer Handbook on UQ*, R. Ghanem, D. Higdon and H. Owhadi (Eds), in press.
- P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa and B. Iooss (2015), *Density modification based reliability sensitivity analysis*, *Journal of Statistical Computation and Simulation*, 85:1200-1223.
- J. Morio et M. Balesdent (2015), *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, A Practical Approach*, Woodhead Publishing.
- J. Fort, T. Klein, N. Rachdi (2013), *New sensitivity analysis subordinated to a contrast* Communication in Statistics : Theory and Methods. In press

[ Roman Sueur; EDF R&D, 6 Quai Watier, 78401 Chatou, France ]  
 [ roman.sueur@edf.fr - ]