Density modification based reliability sensitivity analysis

Paul Lemaître, Ekaterina Sergienko, Aurélie Arnaud, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss

Density modification based reliability sensitivity analysis

P. Lemaître[a b **] E. Sergienko[c d] A. Arnaud[a] N. Bousquet[a] F. Gamboa[d]
B. Iooss[a d]

[a] *EDF R&D, 6 Quai Watier - 78401 Chatou*;
[b] *INRIA Sud-Ouest, 351 cours de la libération - 33405 Talence*;
[c] *IFP EN, 1 avenue de Bois-Préau - 92852 Rueil-Malmaison*;
[d] *Institut de Mathématiques de Toulouse, 118 route de Narbonne - 31062 Toulouse*

October 3, 2012

Abstract

Sensitivity analysis of a numerical model, for instance simulating physical phenomena, is useful to quantify the influence of the inputs on the model responses. This paper proposes a new sensitivity index, based upon the modification of the probability density function (pdf) of the random inputs, when the quantity of interest is a failure probability (probability that a model output exceeds a given threshold). An input is considered influential if the input pdf modification leads to a broad change in the failure probability. These sensitivity indices can be computed using the sole set of simulations that has already been used to estimate the failure probability, thus limiting the number of calls to the numerical model. In the case of a Monte Carlo sample, asymptotical properties of the indices are derived. Based on Kullback-Leibler divergence, several types of input perturbations are introduced. The relevance of this new sensitivity analysis method is analysed through three case studies.

1 Introduction

In the context of structural reliability, computer codes are used in order to assess the safety of industrial systems relying on complex physical phenomena. For instance, an electric operator would like to predict the height of a potential river flood in order to determine the height of a dyke preventing any disaster. In this example, the computer code (simulating the hydraulic model) has some uncertain input variables (flow rate, river length, water height, etc.), that are modelled by random variables. In this paper, the computer code is a "black-box" deterministic numerical model and one of its output is considered. Due to the randomness of the model inputs, this output is a random variable more or less sensitive to the uncertainty of the input variables.

Sensitivity analysis (SA) is a tool used to explore, understand and (partially) validate computer codes. It aims at explaining the outputs regarding the input uncertainties ([13]). The definition of SA differs from fields and authors. We use the "global SA" definition given by Saltelli *et* al. [14] wherein the whole variation range of the inputs is considered. The application of such an approach can be model simplification (by removing irrelevant modelling elements), input variables ranking or research prioritization. There is a wide range of SA techniques, regarding what type of problem the experimenter faces with ([8]). For instance, screening methods are to be applied when there is a large number of inputs, and few models assumptions. For a quantitative point of view, the most popular techniques are variance-based methods, based upon the functional Hoeffding variance decomposition [1] and the so-called Sobol' indices ([14]).

It should be noted that most SA methods focus on real-valued continuous numerical output variables. When the output is a binary value (e.g. when the numerical model returns "faulty system" or "safe system"),

**Corresponding author. Email: paul.lemaitre@edf.fr

SA techniques are underdeveloped. Some basic techniques can be quoted, such as Monte-Carlo filtering ([14]) which consists in measuring differences between a "safe" sample and a "faulty" sample via standard statistical tests. In a different scientific field, the reliability index resulting from the First or Second Order Reliability Methods (FORM/SORM, [9]) can also be used to classify the impact of the inputs on the failure probability. More recent works give methods combining always the two objectives: estimating a failure probability and assessing the influence of the input uncertainty on the failure probability ([11, 12]).

In this paper, a real-valued numerical model denoted by $G : \mathbb{R}^d \to \mathbb{R}$ is considered. This model may further be called the "failure function". In practice, each run of G can be CPU time consuming. We are interested in the event $G(\mathbf{X}) < 0$ (system failure) and in the complementary event $G(\mathbf{X}) \geq 0$ (system safe mode). $\mathbf{X} = (X_1, ..., X_d)^T$ is a d-dimensional continuous random variable whose joint probability density function (pdf) is denoted f. For $i = 1, \cdots, d$, let f_i denotes the distribution of X_i (the marginal pdf). We make the assumption that all components of $\mathbf{X}$ are independent. The quantity of interest is the system failure probability:

$$P \quad = \quad \int 1_{\{G(\mathbf{x})<0\}} f(\mathbf{x}) d\mathbf{x}.$$

The aim of this work is the quantification of the influence of each variable X_i on this probability, by using the same set of calculations that have been used for the estimation of the failure probability.

In most cited works, sensitivity indices for failure probabilities were defined in strong correspondence with a given method of estimation (e.g. [9, 12]), and their interpretation is consequently limited, as stressed in [10]. To answer to genericity concerns expressed by these authors, this article first aims at defining sensivity indices that have more intrinsic relevance (Section 2). Nonetheless, they have to be estimated in practice in function of the method. For simplicity reasons, a classical Monte Carlo framework is considered to estimate P and the indices. It is also useful to derive the theoretical properties of the estimators of the indices. Pursuing the same idea of offering extended tools of sensitivity analysis, Section 3 focuses on generic strategies of input perturbation based upon maximum entropy rules. The behaviour of the indices is examined in Section 4 through numerical simulations in various complexity settings, involving toy examples and a realistic case-study. Comparisons with two reference methods (FORM indices and Sobol' indices) highlight the relevance of the new indices in most situations. The main advantages and remaining issues are finally discussed in the last section of the article. That introduces avenues for future research.

2 Definition, estimation and properties of a sensitivity index

Given a unidimensional input variable X_i with pdf f_i and some perturbation parameter δ lying in a given subset of $\mathbb{R}$, let call $X_{i\delta} \sim f_{i\delta}$ the corresponding perturbed random input. Accordingly, the failure probability becomes

$$P_{i\delta} = \int 1_{\{G(\mathbf{x})<0\}} \frac{f_{i\delta}(x_i)}{f_i(x_i)} f(\mathbf{x}) d\mathbf{x} \tag{1}$$

where x_i is the i^{mboxth} component of the vector $\mathbf{x}$. Independently of the mechanism chosen for the perturbation (see next Section for proposals), a good sensitivity index $S_{i\delta}$ should have intuitive features that make it appealing to reliability engineers and decision-makers. We believe that the following proposal can fulfil these requirements:

$$S_{i\delta} \quad = \quad \left[\frac{P_{i\delta}}{P} - 1\right] \mathbf{1}_{\{P_{i\delta} \geq P\}} + \left[1 - \frac{P}{P_{i\delta}}\right] \mathbf{1}_{\{P_{i\delta} < P\}} \quad = \quad \frac{P_{i\delta} - P}{P \cdot \mathbf{1}_{\{P_{i\delta} \geq P\}} + P_{i\delta} \cdot \mathbf{1}_{\{P_{i\delta} < P\}}} \, .$$

Firstly, $S_{i\delta} = 0$ if $P_{i\delta} = P$, as expected if X_i is a non-influential variable or if δ expresses a negligible perturbation. Secondly, the sign of $S_{i\delta}$ indicates how the perturbation impacts the failure probability qualitatively. It highlights the situations when $P_{i\delta} > P$ amounts to determining if the remaining (*epistemic*) uncertainty on the modelling $X_i \sim f_i$ can increase the failure risk and therefore should be more accurately analysed. Conversely, P can be interpreted as a conservative assessment of the failure probability, robust to perturbations on X_i, if $P_{i\delta} < P$. In such a case, deeper modelling studies on X_i appear less essential. Thirdly,

given its sign the absolute value of $S_{i\delta}$ has simple interpretation and provides a level of the conservatism or non-conservatism induced by the perturbation: a value of $\alpha > 0$ for the index means that $P_{i\delta} = (1+\alpha)P$. If $S_{i\delta} = -\alpha < 0$ then $P_{i\delta} = (1/(1+|\alpha|))P$.

The postulated ability of $S_{i\delta}$ to enlighten the sensitivity of P to input perturbations must be tested in concrete cases, when an estimator $\hat{P}_N$ of P can be computed using an already available design of N numerical experiments. In this paper, N is assumed to be large enough such that statistical estimation stands within the framework of asymptotic theory. Besides, we assume for simplicity a standard Monte Carlo design of experiments, according to which $\hat{P}_N = \sum_{n=1}^{N} 1_{\{G(\mathbf{x}^n)<0\}}/N$ where the $\mathbf{x}^1, \cdots, \mathbf{x}^N$ are independent realisations of X. The strong Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) ensure that for almost all realisations $\hat{P}_N \xrightarrow[N\to\infty]{} P$ and

$$\sqrt{N/[P(1-P)]}(\hat{P}_N - P) \quad \xrightarrow[N\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1).$$

An interest of the Monte Carlo framework is that $P_{i\delta}$ can be consistently estimated without new calls to G, through a "reverse" importance sampling mechanism:

$$\hat{P}_{i\delta N} \quad = \quad \frac{1}{N}\sum_{n=1}^{N} 1_{\{G(\mathbf{x}^n)<0\}} \frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)}.$$

This property holds in the more general case when P is originally estimated by importance sampling rather than simple Monte Carlo, which is more appealing in contexts when G is time-consuming [3, 7]. This generalization is discussed further in the text (Section 5).

Lemma 2.1 *Assume the usual conditions*

(i) $Supp(f_{i\delta}) \subseteq Supp(f_i)$,

(ii) $\displaystyle\int_{Supp(f_i)} \frac{f_{i\delta}^2(x)}{f_i(x)}\, dx < \infty$,

then $\hat{P}_{i\delta N} \xrightarrow[N\to\infty]{} P_{i\delta}$ *and* $\sqrt{N}\sigma_{i\delta N}^{-1}\left(\hat{P}_{i\delta N} - P_{i\delta}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$. *The exact expression of* $\sigma_{i\delta N}^{-1}$ *is given in Appendix A, equation 10. It can be consistently estimated by*

$$\hat{\sigma}_{i\delta N}^2 \quad = \quad \frac{1}{N}\sum_{n=1}^{N} 1_{\{G(\mathbf{x}^n)<0\}} \left(\frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)}\right)^2 - \hat{P}_{i\delta N}^2.$$

The proof of this Lemma is given in Appendix A.1

The asymptotic properties of any estimator of $S_{i\delta}$ will depend on the correlation between $\hat{P}_N$ and $\hat{P}_{i\delta N}$. The next proposition summarizes the features of the joint asymptotic distribution of both estimators.

Proposition 2.1 *Under assumptions (i) and (ii) of Lemma 2.1,*

$$\sqrt{N}\left[\begin{pmatrix} \hat{P}_N \\ \hat{P}_{i\delta N} \end{pmatrix} - \begin{pmatrix} P \\ P_{i\delta} \end{pmatrix}\right] \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_2\left(0, \Sigma_{i\delta}\right)$$

where $\Sigma_{i\delta}$ *is given in Appendix A, equation 11 and can be consistently estimated by*

$$\hat{\Sigma}_{i\delta} \quad = \quad \begin{pmatrix} \hat{P}_N(1-\hat{P}_N) & \hat{P}_{i\delta N}(1-\hat{P}_N) \\ \hat{P}_{i\delta N}(1-\hat{P}_N) & \hat{\sigma}_{i\delta N}^2 \end{pmatrix}.$$

The proof of this Proposition is given in Appendix A.2

Given $(\hat{P}_N, \hat{P}_{i\delta N})$, the plugging estimator for $S_{i\delta}$ is

$$\hat{S}_{\mathrm{i}\delta N} = \left[\frac{\hat{P}_{i\delta N}}{\hat{P}_N} - 1\right] \mathbf{1}_{\left\{\hat{P}_{i\delta N} \geq \hat{P}_N\right\}} + \left[1 - \frac{\hat{P}_N}{\hat{P}_{i\delta N}}\right] \mathbf{1}_{\left\{\hat{P}_{i\delta N} < \hat{P}_N\right\}}. \tag{2}$$

In corollary of Proposition 2.1, applying the continuous-mapping theorem to the continuous function $s(x,y) = \left[\frac{y}{x} - 1\right] \mathbf{1}_{y \geq x} + \left[1 - \frac{x}{y}\right] \mathbf{1}_{y<x}$, $\hat{S}_{\mathrm{i}\delta N}$ converges a.s. to $S_{i\delta}$.

Proposition 2.2 *Assume that assumptions (i) and (ii) of Lemma 2.1 hold and further that $P \neq P_{i\delta}$. We have*

$$\sqrt{N}\left[\hat{S}_{i\delta N} - S_{i\delta}\right] \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, d^T \Sigma d\right) \tag{3}$$

with $d = (\frac{\partial s}{\partial x}(P, P_{i\delta}), \frac{\partial s}{\partial y}(P, P_{i\delta}))^T$ for $x \neq y$, and

$$\begin{aligned} \frac{\partial s}{\partial x}(x,y) &= -y 1_{\{y \geq x\}}/x^2 - 1_{\{y<x\}\frac{1}{y}}, \\ \frac{\partial s}{\partial y}(x,y) &= \frac{1}{x} 1_{\{y \geq x\}} + x 1_{\{y<x\}}/y^2. \end{aligned}$$

The following CLT results from Theorem 3.1 in [16]. Notice that it is also the case when $P = P_{i\delta}$. Indeed, one has for $x^* \neq 0$:

$$\lim_{\substack{y \geq x \\ \binom{x}{y} \to \binom{x^*}{y^*}}} \nabla s(x,y) = \lim_{\substack{y < x \\ \binom{x}{y} \to \binom{x^*}{y^*}}} \nabla s(x,y) = \left(-\frac{1}{x^*}, \frac{1}{x^*}\right)^T.$$

3 Methodologies of input perturbation

Our sensitivity analysis method requires to define a perturbation for each input. In general, and especially in preliminary reliability studies, there is no prior rule allowing to elicit a specialized perturbation for each input variable. When conducting such an analysis, it is advisable to propose one or several fair methodologies for perturbing the inputs.

More precisely, we suggest to define a perturbed input density $f_{i\delta}$ as the closest distribution to the original f_i in the entropy sense and under some constraints of perturbation. Information-theoretical arguments ([5]) led us to choose the Kullback-Leibler (KL) divergence between $f_{i\delta}$ and f_i as a measure of the discrepancy to minimize under those constraints. Recall that between two pdf p and q we have

$$KL(p,q) = \int_{-\infty}^{+\infty} p(y) \log \frac{p(y)}{q(y)} dy \text{ if } \log \frac{p(y)}{q(y)} \in L^1(p(y)dy). \tag{4}$$

Let $i = 1, \cdots, d$, the constraints are linear as functional of the modified density f_{mod}:

$$\int g_k(x_i) f_{\mathrm{mod}}(x_i) dx_i = \delta_{k,i} \ \ (k = 1 \cdots K). \tag{5}$$

Here, for $k = 1, \cdots, K$, g_k are given functions and $\delta_{k,i}$ are given real. These quantities will lead to a perturbation of the original density. The modified density $f_{i\delta}$ considered in our work is:

$$f_{i\delta} = \underset{f_{\mathrm{mod}} | (5) \text{ holds}}{\operatorname{argmin}} KL(f_{\mathrm{mod}}, f_i) \tag{6}$$

and the result takes an explicit form ([6]) given in the following proposition.

Proposition 3.1 *Let us define, for* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_K)^T \in \mathbb{R}^K$,

$$\psi_i(\boldsymbol{\lambda}) = \log \int f_i(x) \exp\left[\sum_{k=1}^{K} \lambda_k g_k(x)\right] dx \ , \tag{7}$$

where the last integral can be finite or infinite (in this last case $\psi_i(\boldsymbol{\lambda}) = +\infty$*). Further, set dom* $\psi_i = \{\boldsymbol{\lambda} \in \mathbb{R}^K | \psi_i(\boldsymbol{\lambda}) < +\infty\}$*. Assume that there exists at least one pdf* f_m *satisfying (5) and that dom* ψ_i *is an open set. Then, there exists a unique* $\boldsymbol{\lambda}^*$ *such that the solution of the minimisation problem (6) is*

$$f_{i\delta}(x_i) = f_i(x_i) \exp\left[\sum_{k=1}^{K} \lambda_k^* g_k(x_i) - \psi_i(\boldsymbol{\lambda^*})\right].$$

The theoretical technique to compute $\boldsymbol{\lambda}$ is provided in appendix B. Hereby are presented two kinds of perturbations used further on.

Mean twisting The first moment is often used to parametrize a distribution. Thus the first perturbation presented here is a mean shift, that is expressed with a single constraint:

$$\int x_i f_{\text{mod}}(x_i) dx_i = \delta_i \ . \tag{8}$$

In term of SA, this perturbation should be used when the user wants to understand the sensitivity of the inputs to a mean shift - that is to say "what if the mean of input X_i were δ_i instead of $\mathbb{E}[X_i]$".

Proposition 3.2 *Considering the constraint (8), under the assumptions of Proposition 3.1 the expression of the optimal perturbed density is*

$$f_{i\delta_i}(x_i) = \exp(\lambda^* x_i - \psi_i(\lambda^*)) f_i(x_i)$$

where λ^* *is such that equation (8) holds.*

It can also be noted that equation (7) becomes

$$\psi_i(\lambda) = \log \int f_i(x_i) \exp(\lambda x_i) dx_i = \log\left(M_{X_i}(\lambda)\right)$$

where $M_{X_i}(u)$ is the moment generating function (mgf) of the $i-$th input. With this notation, λ^* is such that

$$\int x_i \exp\left(\lambda^* x_i - \log\left(M_{X_i}(\lambda^*)\right)\right) f_i(x_i) dx_i = \delta_i \ ,$$

which leads to

$$\int x_i \exp\left(\lambda^* x_i\right) f_i(x_i) dx = \delta_i M_{X_i}(\lambda^*) \ .$$

This equation can be simplified to:

$$\frac{M'_{X_i}(\lambda^*)}{M_{X_i}(\lambda^*)} = \delta_i \ .$$

This equation may be easy to solve when one has the expression of the mgf of the input X_i and of its derivative.

Variance twisting In some cases, the mean of an input may not be the main source of uncertainty, but rather the second moment. This case may be treated considering a couple of constraints. The perturbation presented is a variance shift, therefore the set of constraints is:

$$\begin{cases} \int x_i f_{\text{mod}}(x_i) dx_i = \mathbb{E}[X_i] \ , \\ \int x_i^2 f_{\text{mod}}(x_i) dx_i = V_{\text{per},i} + \mathbb{E}[X_i]^2 \ . \end{cases} \tag{9}$$

The perturbed distribution has the same expectation $\mathbb{E}[X_i]$ as the original one and a perturbed variance $V_{\text{per},i} = \text{Var} X_i \pm \delta_i$.

Proposition 3.3 *Under the assumptions of Proposition 3.1, for the constraint (9), the expression of the optimal perturbed density is:*

$$f_{i\delta_i}(x_i) = \exp(\lambda_1^* x + \lambda_2^* x^2 - \psi_i(\boldsymbol{\lambda}^*)) f_i(x_i)$$

where λ_1^* and λ_2^* are so that equation (9) holds.

Proposition 3.4 *Let consider that the original random variable X_i is distributed according to a Natural Exponential Family (NEF). Recalls that a NEF's pdf is of the form:*

$$f_{i,\theta}(x_i) = b(x_i) \exp\left[x_i\theta - \eta(\theta)\right]$$

where θ is a parameter from a parametric space Θ, $b(.)$ is a function that depends only of x_i and

$$\eta(\theta) = \log \int b(x) \exp\left[x_i\theta\right] dx_i$$

is the cumulant distribution function. Considering the assumptions of Proposition 3.1, then it is straightforward by theorem 3.1 in [6] that optimal pdfs proposed respectively in Proposition 3.2 and Proposition 3.3 are also distributed according to a NEF. The details of computation are given for a mean shift and a variance shift in Appendix D.

4 Numerical experiments

In this Section, the methodology is tested on two academic cases and a more realistic industrial code. The new indices are compared to the results of two reference methods, FORM indices (or Importance Factors, IF) and Sobol' indices (SI). Both are computed using the methodologies given in [9] and [15], respectively. To assess the reproducibility of the estimation of the SI, a sample of 10^5 points is used, and 50 replications are made. Thus all the estimations of the SI are the mean of the obtained values and the coefficient of variation (CV) of the index is provided. One should notice that the SI are applied on the indicator of the failure function $1_{\{G(\mathbf{x})<0\}}$. Following the definition of IF and SI, those indices lies in $[0, 1]$.

4.1 Hyperplane failure surface

For the first example, $\mathbf{X}$ is set to be a $4-$dimensional vector, with $d = 4$ independent marginal distributions normally distributed with parameters 0 and 1. Therefore $f_{X_i} \sim \mathcal{N}(0, 1)$ for $i = 1, .., 4$. The failure function is defined as:

$$G(\mathbf{X}) = k - \sum_{i=1}^{4} a_i X_i$$

where k and $\mathbf{a} = (a_1, a_2, a_3, a_4)$ are the parameters of the model. For this numerical example, parameters are set with values $k = 16$ and $\mathbf{a} = (1, -6, 4, 0)$. An explicit expression for P can be given since $G(\mathbf{X})$ behaves like a Gaussian distribution with mean k and standard deviation $\sqrt{\sum_{i=1}^{4} a_i^2}$. Therefore:

$$P = \phi\left(-k/\sqrt{\sum_{i=1}^{4} a_i^2}\right) = \phi\left(\frac{-16}{\sqrt{53}}\right) \simeq 0.014$$

where $\phi(.)$ is the standard normal cumulative distribution function.

It is expected that the influence of X_i on P uniquely depends on $|a_i|$. The greater the absolute value of the coefficient is, the bigger the expected influence is. The aim of choosing one non-influential (dummy) variable X_4 (because $a_4 = 0$) is to assess if the SA methods can identify this variable as non-influential on the failure probability.

4.1.1 FORM

In this ideal hyperplane failure surface case, FORM performs well as expected [9] by providing an approximated value $\hat{P}_{FORM} = 0.01398$. The importance factors, given in Table 1, provide an accurate variable ranking for the failure function, given the a_i factors.

Variable	X_1	X_2	X_3	X_4
Importance factor	0.018	0.679	0.302	0

Table 1: Importance factors for hyperplane function

4.1.2 Sobol' indices

The first-order and total indices are displayed in Table 2. The interpretation of the results is that X_2 and X_3 concentrate most of the variance of the indicator function. At first order, 25% of its variance is explained by X_2 without any interaction. It should be noted that the total index for X_4 is null, assessing that this variable does not impact the failure probability. The CV of the total indices estimators are small, meaning that this method is reproducible and that 10^5 points are enough to estimate in an efficient way the indices S_{Ti}. On the other hand, some CV values for low mean first order indices are quite high. The conclusion of this result is that the method correctly estimates high indices but estimates poorly the indices close to 0. On the other hand, the relevant information is that the index is close to 0. Thus this situation may not be a problem.

Sobol' Index	S_1	S_2	S_3	S_4	S_{T1}	S_{T2}	S_{T3}	S_{T4}
Mean	0.0017	0.2575	0.0544	$9.45.10^{-5}$	0.1984	0.9397	0.7256	0
C.o.v.	1.5854	0.04826	0.1336	27.4	0.012	0.0069	0.013	0

Table 2: Sobol' indices for hyperplane function

4.1.3 Density modification based reliability indices

The method presented throughout this article is applied on the hyperplane function. As explained in Section 3, several ways to perturb the input distributions exist. For this case, we choose to apply first a mean twisting, then a variance twisting with fixed mean. A simple calculus gives that the perturbed pdf are Gaussian, respectively with the constraint mean and variance 1 for the mean twisting perturbation (see Table **??**), and with mean 0 and the constraint variance for the variance twisting perturbation. Thus, the MC estimation gives $\hat{P} = 0.01446$. For the mean twisting (see (8)), the variation range chosen for δ is from -1 to 1 with 40 points, reminding that $\delta = 0$ cannot be considered as a perturbation. For the variance twisting (see (9)), the variation range chosen for V_{per} is from 1/20 to 3 with 28 points, where $V_{\text{per}} = 1$ is not a perturbation. The estimated indices are plotted respectively in Figure 1 for mean twisting and in Figure 2 for variance twisting. 95% confidence intervals are plotted around the indices.

Mean perturbation indices The indices $\widehat{S_{i\delta}}$ behave in a monotonic way given the importance of the perturbation. The slope at the origin is directly related to the value of a_i. For influential variables (X_2 and X_3), the increasing or the decreasing is faster than linear, whereas the curve seems linear for the slightly influential variable (X_1). Modifying the mean of amplitude δ positive slightly rises the failure probability for X_1, highly decreases it for X_2 and increases it for X_3 (Figure 1). The effects are reversed with similar amplitude for negative δ. It can be seen that X_4 has no impact on the failure probability for any perturbation. Those results are consistent with the expression of the failure function. One can see that the confidence intervals (CI) associated to X_2 and X_3 are fairly well separated, except for the small absolute value of δ. On the other hand, the CI associated to X_1 and X_4 are not separated until absolute value of δ higher than 0.6. The conclusion of the observation of the CI is that one cannot differentiate the impact of variable X_1 and X_4 unless a broad change of the mean occurs.

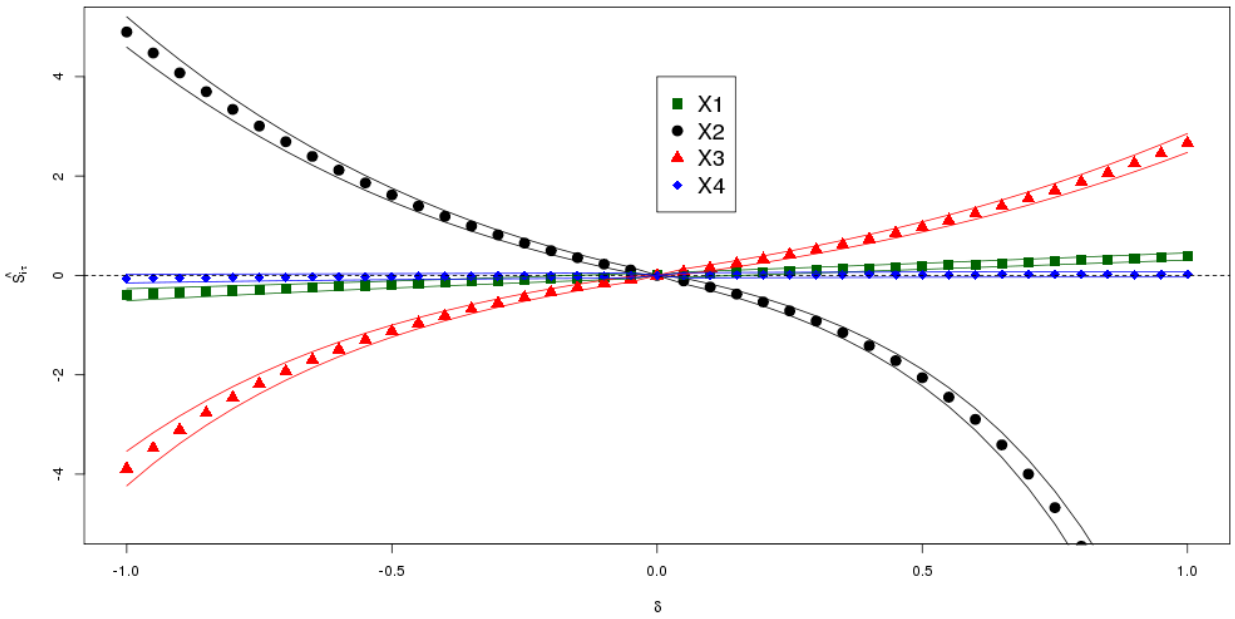

Figure 1: Estimated indices $\widehat{S_{i\delta}}$ for hyperplane function with a mean twisting

Variance perturbation indices Increasing the variance of input X_2 and X_3 increases the failure probability, whereas it is the opposite when decreasing the variance (Figure 2). Modifying the variance of X_1 and X_4 have no effect on the failure probability. The increasing of the indices is linear for X_2 and X_3, and the decreasing of the indices is faster than linear, especially for X_2. Considering the CI, one can see that they are well separated for variable X_2 and X_3, assessing the relative importance of these variables. On the other hand, as the CI associated to X_1 and X_4 are not separated and contain 0.

4.2 Thresholded Ishigami function

The Ishigami function is a common test case in SA since it has a complex expression, with interactions between the variables. A modified version of the Ishigami function will be considered in this paper. One has:

$$G(\mathbf{X}) = \sin(X_1) + 7\sin(X_2)^2 + 0.1X_3^4\sin(X_1) + 7$$

where $\mathbf{X}$ is a $3-$dimensional vector of independent marginals uniformly distributed on $[-\pi, \pi]$. In Figure 3, the failure points (where $G(\mathbf{x}) < 0$) are plotted in a 3-d scatterplot.

There are 614 failure points on a MC sample of 10^5 points therefore the failure probability here is roughly $\hat{P} = 6.14.10^{-3}$. The complex repartition of the failure points can be noticed. Those points lay in a zone defined by the negative values of X_1, the extremal and mean values of X_2 (around $-\pi$, 0 and π), and the extremal values of X_3 (around $-\pi$ and π).

4.2.1 FORM

The algorithm FORM converges to an incoherent design point $(6.03, 0.1, 0)$ in 50 function calls, giving an approximate probability of $\hat{P}_{FORM} = 0.54$. The importance factors are displayed in Table 3. The bad performance of FORM is expected given that the failure domain consists in six separate domains and that the function is highly oscillant, leading to optimization difficulties. The design point is aberrant, thus the FORM results of SA are incorrect.

4.2.2 Sobol' indices

The first-order and total indices are displayed in Table 4. The small values of first order indices show that no variable has impact on the variance of the indicator of failure on its own. The three total indices have

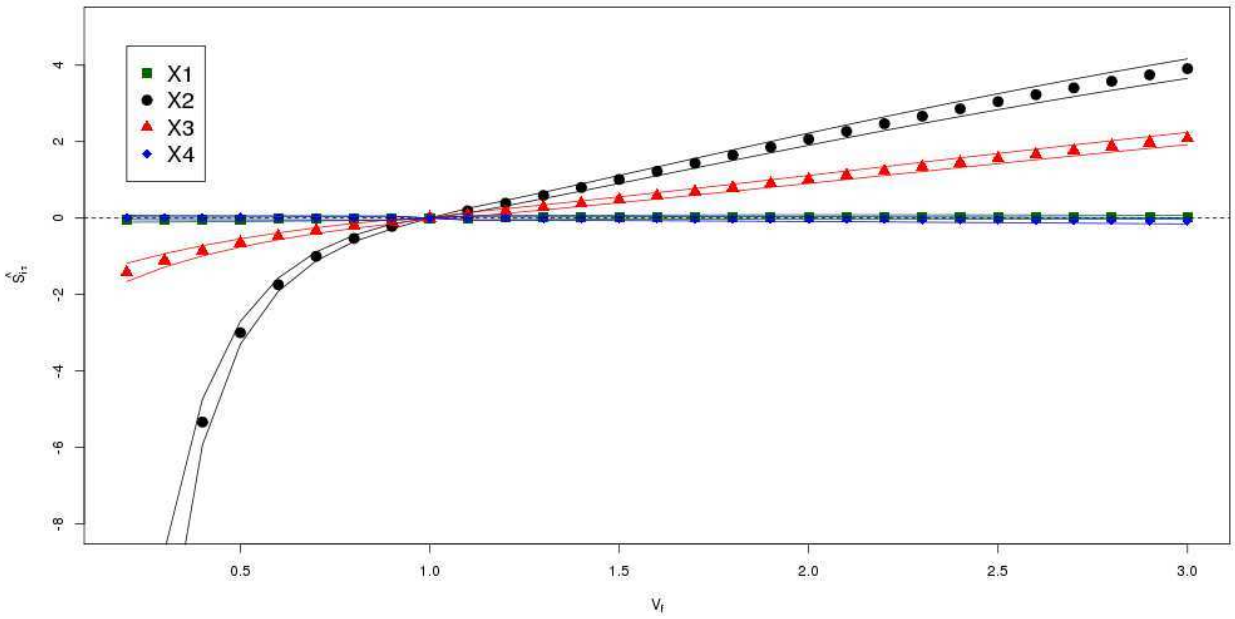

Figure 2: Estimated indices $\widehat{S_{i,V_{\text{per}}}}$ for hyperplane function with a variance twisting

Variable	X_1	X_2	X_3
Importance factor	$1e^{-17}$	1	0

Table 3: Importance factors for Ishigami function

relatively high and similar values. This states that all the variables highly interact with each other to cause system failure. The SI method is thus non-discriminant in this case. The low CV shows that the method is reproducible.

Sobol' Index	S_1	S_2	S_3	S_{T1}	S_{T2}	S_{T3}
Mean	0.0234	0.0099	0.0667	0.8158	0.6758	0.9299
C.o.v.	0.0072	0.0051	0.0095	0.0156	0.0216	0.0094

Table 4: Sobol' indices for Ishigami function

4.2.3 Density modification based reliability indices

The method presented throughout this article is applied on the thresholded Ishigami function. As for the hyperplane test case, a mean twisting and a variance twisting are applied. The modified distribution when a mean shift is applied on a uniform distribution is given in Table **??**. The modified pdf when shifting the variance and keeping the same expectation is proportional to a truncated Gaussian when decreasing the variance. When increasing the variance, the perturbed distribution is a symmetrical distribution with 2 modes close to the endpoints of the support. As previously, the same MC sample of size 10^5 (also used to produce Figure 3) is used to estimate the indices with both perturbations. For the mean twisting (see (8)), the variation range chosen for δ is -3 to 3 with 60 points - numerical consideration forbidding to choose a shifted mean closer to the endpoints. For variance twisting, the variation range chosen for V_{per} is 1 to 5 with 40 points. Let us recall that the original variance is $\text{Var}[X_i] = \pi^2/3 \simeq 3.29$. The estimated indices are plotted respectively in Figure 4 for mean twisting and in Figure 5 for variance twisting.

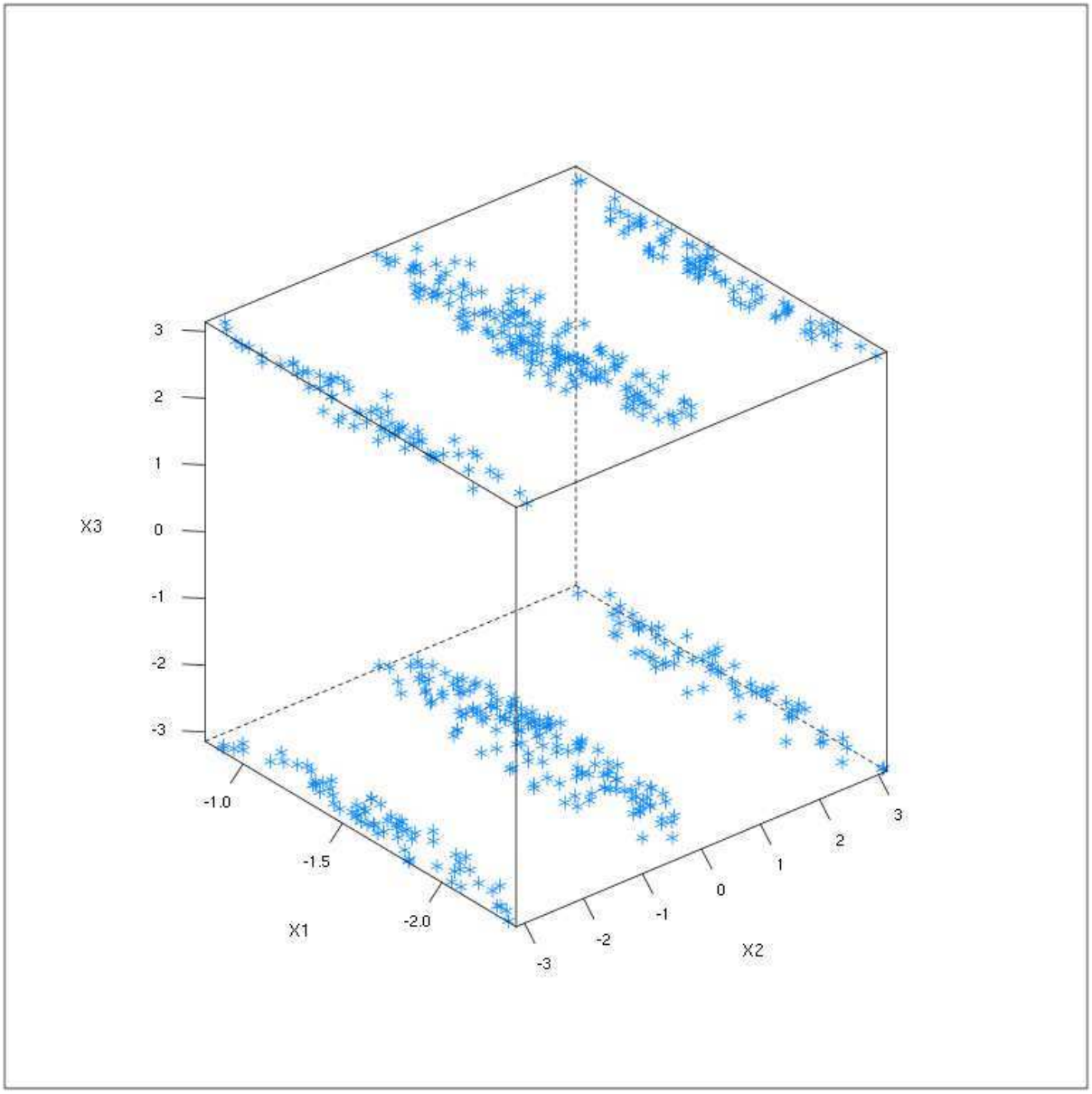

Figure 3: Ishigami failure points from a MC sample

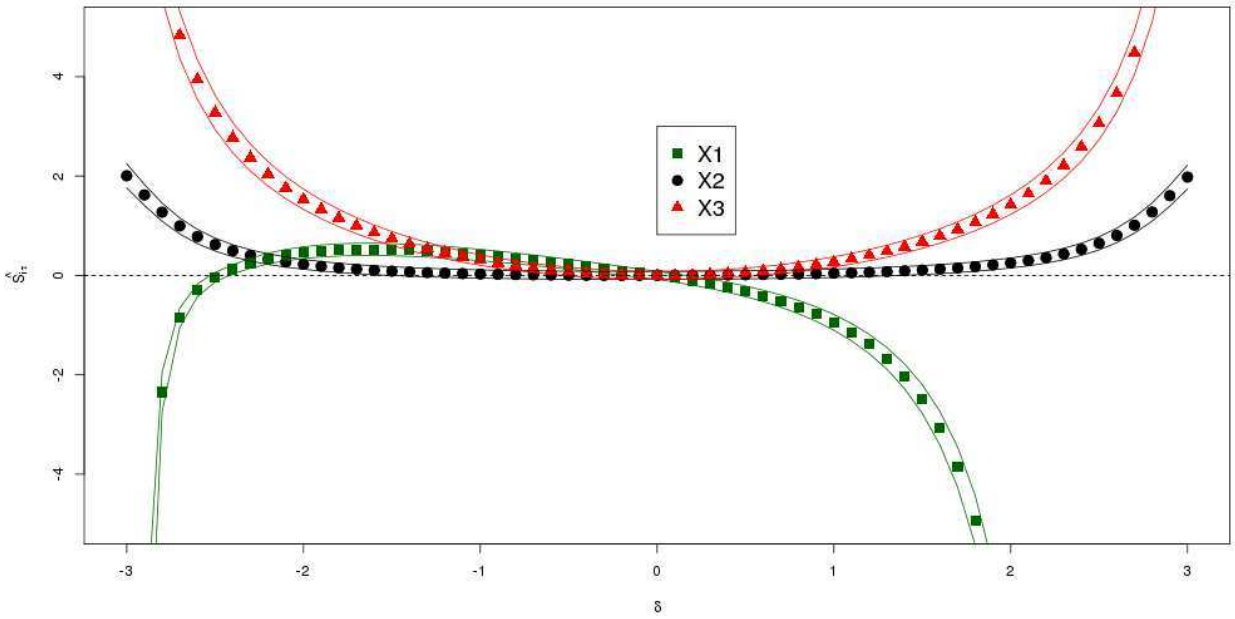

Figure 4: Estimated indices $\widehat{S_{i\delta}}$ for the thresholded Ishigami function with a mean twisting

Mean perturbation indices A perturbation of the mean for X_2 and X_3 will increase the failure probability, though the impact for the same mean perturbation is stronger for X_3 ($\widehat{S_{3,-3}}$ and $\widehat{S_{3,3}}$ approximately equal respectively 9.5 and 10, Figure 4). On the other hand, the indices concerning X_1 show that a mean shift between -1 and -2 increases the failure probability, whereas an augmentation of the mean or a large diminution strongly diminishes the failure probability ($\widehat{S_{1,3}}$ approximatively equals -7.10^{11}). Therefore, Figure 4 leads to two conclusions. First, the failure probability can be strongly reduced when shifting the mean of the first X_1 (this is also provided by Figure 3 wherein all failure points have a negative value of X_1). Second, any change in the mean for X_2 or X_3 will lead to an increase of the failure probability. The CI are well separated, except in the -1 to 1 zone. One can notice that the CI associated to X_2 contains 0 between values of δ from -1.5 to 1.5, thus the associated indices might be null in these case. This has to be taken into account when assessing the relative importance of X_2.

Variance perturbation indices Figure 5 (upper) shows that a change in the variance has little effect on X_2 and X_1, though the change is of opposite effect on the failure probability. However, considering that the indices $\widehat{S_{2,V_{\text{per},i}}}$ and $\widehat{S_{1,V_{\text{per},i}}}$ lie between -0.4 and 0.4, one can conclude that the variance of theses variables are not of great influence on the failure probability. On the other hand, Figure 5 (lower) shows that any reduction of $\text{Var}[X_3]$ strongly decreases the failure probability, and that an increase of the variance slightly increases the failure probability. This is relevant with the expression of the failure surface, as X_3 is fourth powered and multiplied by the sinus of X_1. A variance decrease as formulated gives a distribution concentrated around 0. Decreasing $\text{Var}[X_3]$ shrinks the concerned term in $G(\mathbf{X})$. Therefore it reduces the failure probability. The CI associated to X_3 are broadly separated from the others. On the other hand, the CI associated to X_1 and X_2 overlap when the fixed variance goes from 1.5. to 4. It is thus not possible to conclude with certainty on the difference of impact of those variables.

4.3 Industrial case : flood case

The goal of this test case is to assess the risk of a flood over a dyke for the safety of industrial installations. This comes down to model the height of a flood. Given the uncertainty upon numerous physical parameters, the uncertainty approach is used and unknown parameters are modelled by random variables. From a simplification of the Saint-Venant equation, a flood risk model is obtained. The quantity of interest is the difference between the height of the dyke and the height of water. If this quantity is negative, the installation

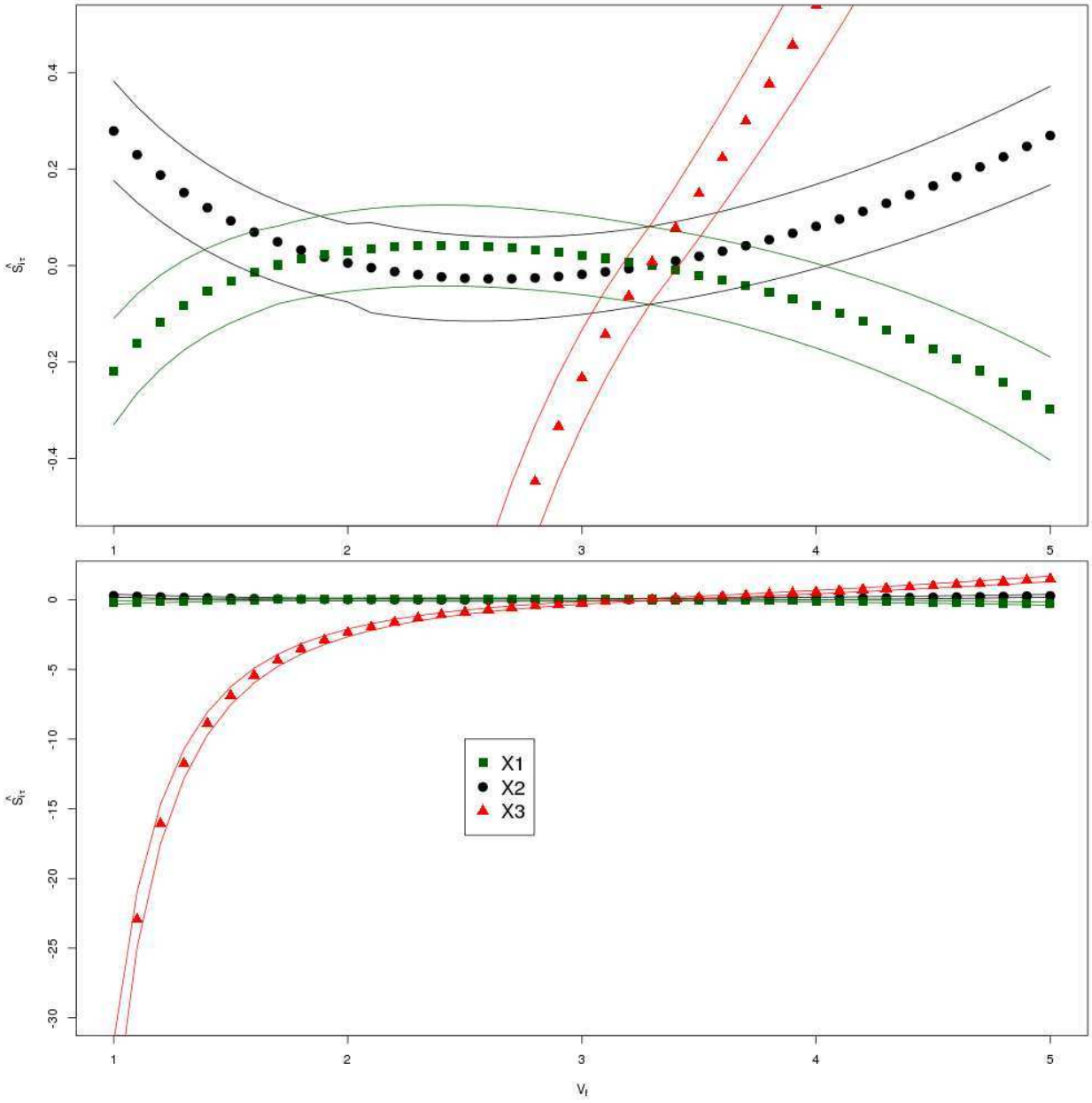

Figure 5: Estimated indices $\widehat{S_{i,V_{\text{per}}}}$ for the thresholded Ishigami function with a variance twisting. The upper figure is a zoom where the $\widehat{S_{i,V_{\text{per}}}}$ axis lies into $[-0.5, 0.5]$. The lower figure shows almost the whole range variation for $\widehat{S_{i,V_{\text{per}}}}$.

is flooded; this is the failure event that is considered. Several quantities will be denoted as follows: Q the flow rate, L the watercourse section length studied, B the watercourse width, K_s the watercourse bed friction coefficient (also called Strickler coefficient), Z_m and Z_v respectively the upstream and downstream bottom watercourse height above sea level and H_d the dyke height measured from the bottom of the watercourse bed. The water height model is expressed as:

$$H = \left(\frac{Q}{K_s B \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{\frac{3}{5}}.$$

Therefore the following quantity is considered:

$$G = H_d - (Z_v + H).$$

Among the model inputs, the choice is made that the following variables are known precisely: $L = 5000$ (m), $B = 300$ (m), $H_d = 58$ (m), and the following are considered to be random. Q ($\mathrm{m}^3.\mathrm{s}^{-1}$) follows a positively truncated Gumbel distribution of parameters $a = 1013$ and $b = 558$ with a minimum value of 0. K_s ($\mathrm{m}^{1/3}\mathrm{s}^{-1}$) follows a truncated Gaussian distribution of parameters $\mu = 30$ and $\sigma = 7.5$, with a minimum value of 1. Z_v (m) follows a triangular distribution with minimum 49, mode 50 and maximum 51. Z_m (m) follows a triangular distribution with minimum 54, mode 55 and maximum 56.

4.3.1 FORM

The algorithm FORM converges to a design point $(1.72, -2.70, 0.55, -0.18)$ in 52 function calls, giving an approximate probability of $\hat{P}_{FORM} = 5.8.10^{-4}$. The importance factors are displayed in Table 5.

Variable	Q	K_s	Z_v	Z_m
Importance factor	0.246	0.725	0.026	0.003

Table 5: Importance factors for flood case

FORM assesses that K_s is of extremely high influence, followed by Q that is of medium influence. Z_v has a very weak influence and Z_m is negligible. It can be noticed that the estimated failure probability is twice as small as the one estimated with crude MC, but remains in the same order of magnitude.

4.3.2 Sobol' indices

The first-order and total indices are displayed in Table 6. It can be seen that the estimates of some indices are negative despite the fact that Sobol indices are theoretically positive. The estimation can indeed produce negative results for values close to 0.

Sobol Index	S_Q	S_{K_s}	S_{Z_v}	S_{Z_m}	S_{TQ}	S_{TK_s}	S_{TZ_v}	S_{TZ_m}
Mean	0.0169	0.2402	-7.10^{-5}	-5.10^{-4}	0.7447	0.9782	0.2684	0.1062
C.o.v.	0.0122	0.0577	0.0029	0.0023	0.0553	0.0137	0.0516	0.0389

Table 6: Sobol' indices for flood case

Considering the first order indices, Z_v and Z_m are of null influence on their own. Q is considered to have a minimal influence (1% of the variance of the indicator function) by itself, and K_s explains 24% of the variance on its own. When considering the total indices, it can be noticed that both Z_v and Z_m have a weak impact on the failure probability. On the other hand, Q has a major influence on the failure probability. K_s total index is close to one, therefore K_s explains (with or without any interaction with other variables) almost all the variance of the failure function.

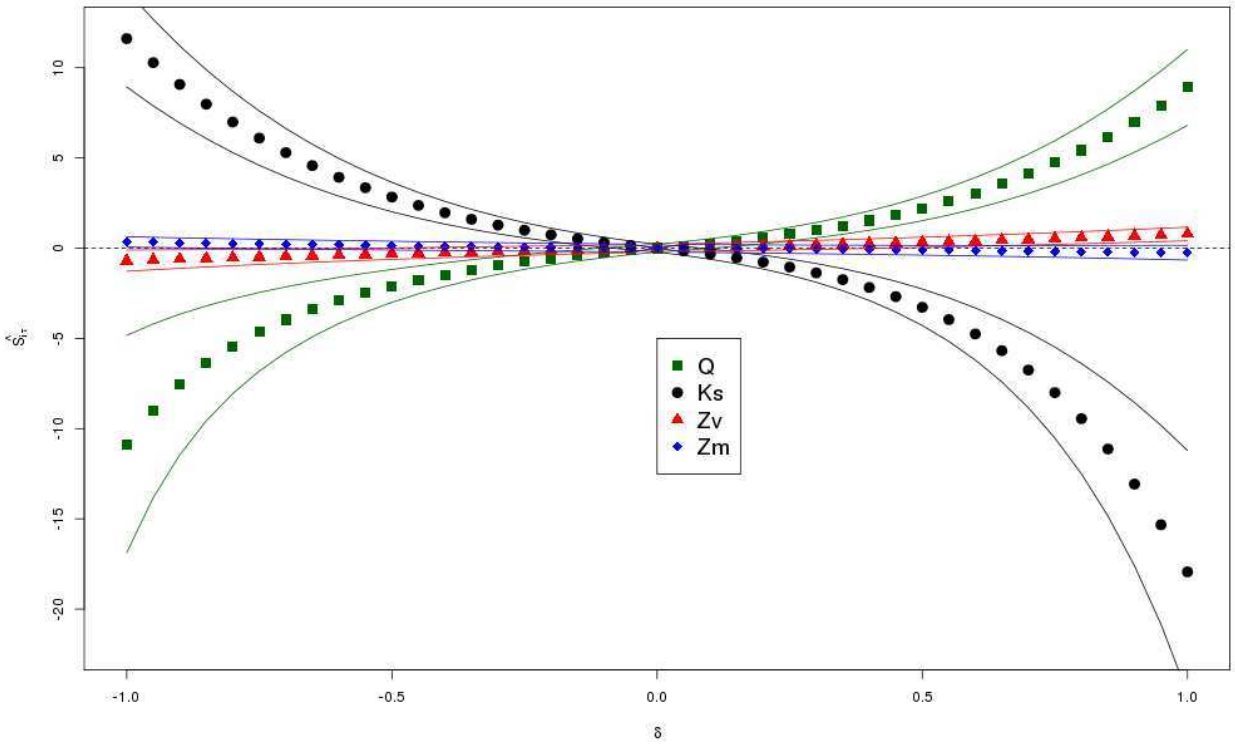

Figure 6: Estimated indices $\widehat{S_{i\delta}}$ for flood case with a mean twisting

4.3.3 Density modification based reliability indices

The method presented throughout this article is applied on the flood case. Only the mean twisting will be applied here. The modified pdf are given in Table in Table **??**. One can notice that the different inputs follow various distributions (unlike the other examples), thus the question of "equivalent" perturbation arises.It will be discussed further in Section 5. Here the choice has been made to shift the mean relatively to the standard deviation, hence including the spread of the various inputs in their respective perturbation. So for any input, the original distribution is twisted so that the perturbed distribution's mean is the original's one plus δ times its standard deviation, δ going from -1 to 1 with 40 points. The 10^5 MC sample gives an estimation of the failure probability $\hat{P} = 8.6.10^{-4}$.

Figure 6 assesses that an increasing of the mean of the inputs increases the failure probability slightly for Z_v, strongly for Q, and diminishes it slightly for Z_m and strongly for K_s. This goes the opposite way when decreasing the mean. In terms of absolute modification, K_s and Q are of same magnitude, even if K_s has a slightly stronger impact. On the other hand, the effects of mean perturbation on Z_m and Z_v are negligible. The CI associated to Q and K_s are well separated from the others, except in a $\delta = -.3$ to .3 zone. The CI associated to Z_v and Z_m overlap, thus even though the indices seem to have different value, it is not possible to conclude with certainty about the influence of those variables.

5 Discussion

5.1 Conclusion on the method

The method presented in this paper gives interesting complementary information in addition of traditional SA methods applied to a reliability problem. Additionally, it has two advantages:

- The ability for the user to set the most adapted constraints considering his/her problem,

- The MC framework allowing to use previously done function calls, thus limiting the CPU cost of the SA, and allowing the user to test several perturbations.

5.2 Equivalent perturbation

The question of "equivalent" perturbation arises from cases where all inputs are not identically distributed. Indeed, problems may emerge when some inputs are defined on infinite intervals and when other inputs are defined on finite intervals (such as uniform distributions). Consider a two-dimensional model with one Gaussian distribution and one uniform distribution as inputs. Thus, a mean shift will be a translation for the first input, whereas it will lead to a Dirac distribution in one endpoint for the other input. Hence, a mean shift cannot be considered as an "equivalent" perturbation. One could think of a "relative mean shift", which seems a fairly good idea. But let one consider a model with two Gaussian inputs of equal variance 1 and of mean respectively 1 and 10000. Then, a relative mean shift of 10% will result in Gaussian distributions with mean respectively 1.1 and 11000, and still variance 1. This counter-example shows that relative mean shift might not be an adequate perturbation in terms of "equivalence".

5.3 Further work

Two main avenues are of interest:

- To adapt the estimator of the indices $S_{i\delta}$, in term of variance reduction and of number of function calls. Further work will be made with importance sampling methods, and possibly subset methods. The use of sequential methods [4] may also be tested,
- To find a way to perturb "equivalently" several distributions of different natures. A perturbation that is not based upon a moment constraint but rather of an entropy constraint might be proposed. The differential entropy of a distribution can be seen as a quantification of uncertainty [2]. Thus an example of (non-linear) constraint on the entropy can be:

$$-\int f_{X_{i\delta}}(x)\log f_{X_{i\delta}}(x)dx = -\delta\int f_{X_i}(x)\log f_{X_i}(x)dx.$$

  Yet further computations have to be made to obtain a tractable solution of the KL minimization problem under the above constraint.

Acknoweldgements

Part of this work has been backed by French National Research Agency (ANR) through COSINUS program (project COSTA BRAVA noANR-09-COSI-015). We thank Dr. Daniel Busby (IFP EN) for several discussions. We also thank Emmanuel Remy (EDF R&D) for proofreading.

References

[1] A. Antoniadis. Analysis of variance on function spaces. *Math. Operationsforsch. Statist. Ser. Statist.*, 15(1):59–71, 1984.

[2] B. Auder and B. Iooss. Global sensitivity analysis based on entropy. In *Proceedings of the ESREL 2008 Conference*, pages 2107–2115. CRC Press, 2008.

[3] R.J. Beckman and M.D McKay. Monte-Carlo estimation under different distributions using the same simulation. *Technometrics*, 29(2):153–160, 1987.

[4] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):1–21, 2011.

[5] T.M. Cover and J.A. Thomas. Elements of information theory 2nd edition (Wiley series in telecommunications and signal processing). 2006.

[6] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[7] T.C. Hesterberg. Estimates and confidence intervals for importance sampling sensitivity analysis. *Mathematical and Computer Modelling*, 23(8):79–85, 1996.

[8] Bertrand Iooss. Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1):1–23, 2011.

[9] M. Lemaire, A. Chateauneuf, and J.C. Mitteau. *Structural reliability*. Wiley Online Library, 2009.

[10] P. Lemaître and A. Arnaud. Hiérarchisation des sources d'incertitudes vis à vis d'une probabilité de dépassement de seuil - une méthode basée sur la pondération des lois. In *Proceedings des 43 èmes Journées de Statistique*, Tunis, Tunisia, June 2011.

[11] J. Morio. Influence of input PDF parameters of a model on a failure probability estimation. *Simulation Modelling Practice and Theory*, 19(10):2244–2255, 2011.

[12] M. Munoz Zuniga, J. Garnier, E. Remy, and E. de Rocquigny. Adaptive directional stratification for controlled estimation of the probability of a rare event. *Reliability Engineering & System Safety*, 92(12):1691–1712, 2011.

[13] A. Saltelli. Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3):579–590, 2002.

[14] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. Sensitivity analysis in practice: A guide to assessing scientific models. 2004. *Chichester, England: John Wiley & Sons*, 46556:48090–9055.

[15] I.M. Sobol. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1(4):407–414, 1993.

[16] A.W. Van der Vaart. *Asymptotic statistics*. Number 3. Cambridge Univ Pr, 2000.

Appendices

A Proofs

A.1 Proof of Lemma 2.1

Under assumption **(i)**, we have

$$\int_{\mathrm{Supp}(f_{i\delta})} 1_{\{G(\mathbf{x})<0\}} \frac{f_{i\delta}(x_i)}{f_i(x_i)} f(\mathbf{x})\ d\mathbf{x} \leq \int_{\mathrm{Supp}(f_{i\delta})} f_{i\delta}(x_i)\ dx_i \ = \ 1.$$

So that, the strong LLN may be applied to $\hat{P}_{i\delta N}$. Defining

$$\sigma^2_{i\delta} = \mathrm{Var}\left[1_{\{G(\mathbf{X})<0\}} \frac{f_{i\delta}(X_i)}{f_i(X_i)}\right], \tag{10}$$

one has

$$\sigma^2_{i\delta} \ = \ \int_{\mathrm{Supp}(f_i)} 1_{\{G(\mathbf{x})<0\}} \frac{f^2_{i\delta}(x_i)}{f_i(x_i)} \prod_{j\neq i}^{d} f_j(x_j)\ d\mathbf{x} - P^2_{i\delta} \ < \ \infty \quad \text{under Condition } \textbf{(ii)}.$$

Therefore the CLT applies:

$$\sqrt{N}\sigma^{-1}_{i\delta}\left(\hat{P}_{i\delta N} - P_{i\delta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\ .$$

Under assumption **(ii)**, the strong LLN applies to $\hat{\sigma}^2_{i\delta N}$. So that, the final result is straightforward using Slutsky's lemma.

A.2 Proof of Proposition 2.1

First, note that

$$\begin{aligned}
\mathbb{E}\left[\widehat{P}\widehat{P_{i\delta}}\right] - PP_{i\delta} &= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{n=1}^{N} 1_{\{G(\mathbf{x}^n)<0\}}\right)\left(\sum_{n=1}^{N} 1_{\{G(\mathbf{x}^n)<0\}}\frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)}\right)\right] - PP_{i\delta} \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\left[1_{\{G(\mathbf{x}^n)<0\}}\right]^2\frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)} + \sum_{n=1}^{N}\sum_{j\neq i}^{N} 1_{\{G(\mathbf{x}^n)<0\}}1_{\{G(\mathbf{x}^j)<0\}}\frac{f_{i\delta}(x_i^j)}{f_i(x_i^j)}\right] \\
&\quad -PP_{i\delta} \\
&= \frac{1}{N^2}\left[NP_{i\delta} + N\left(N-1\right)PP_{i\delta}\right] - PP_{i\delta} \\
&= \frac{1}{N}\left(P_{i\delta} - PP_{i\delta}\right) \ .
\end{aligned}$$

Assuming the conditions under which Lemma 1 is true, the bivariate CLT follows with

$$\Sigma_{i\delta} = \begin{pmatrix} P(1-P) & P_{i\delta}(1-P) \\ P_{i\delta}(1-P) & \sigma_{i\delta}^2 \end{pmatrix} .$$

Each term of this matrix can be consistently estimated, using the results in Lemma 1 and Slutsky's lemma.

B Computation of Lagrange multipliers

Let H be the Lagrange function:

$$H(\boldsymbol{\lambda}) = \psi_i(\boldsymbol{\lambda}) - \sum_{k=1}^{K}\lambda_k\delta_k.$$

Thus, using the results of [6], we have

$$\boldsymbol{\lambda}^* = \arg\min H(\boldsymbol{\lambda}).$$

The expression of the gradient of H with respect to the j^{th} variable is

$$\nabla_j H(\boldsymbol{\lambda}) = \frac{\int g_j(x)f_i(x)\exp(\sum_{k=1}^{K}\lambda_k g_k(x))dx}{\exp\psi_i(\boldsymbol{\lambda})} - \delta_j.$$

In the same way, the expression of the second derivative of H with respect to the h^{th} and the j^{th} variables is

$$\begin{aligned}
D_{hj}H(\boldsymbol{\lambda}) &= \frac{\int g_h(x)g_j(x)f_i(x)\exp(\sum_{k=1}^{K}\lambda_k g_k(x))dx}{\exp\psi_i(\boldsymbol{\lambda})} \\
&- \frac{\int g_j(x)f_i(x)\exp(\sum_{k=1}^{K}\lambda_k g_k(x))dx}{\exp\psi_i(\boldsymbol{\lambda})}\frac{\int g_h(x)f_i(x)\exp(\sum_{k=1}^{K}\lambda_k g_k(x))dx}{\exp\psi_i(\boldsymbol{\lambda})}.
\end{aligned}$$

This method has been used in this paper for computing the optimal vector $\boldsymbol{\lambda}^*$ when a variance twisting was applied. The integral were evaluated with Simpson's rule.

C Numerical trick to work with truncated distribution

In the case where a mean twisting is considered on a left truncated distribution, here is presented a tip that can help to compute $\boldsymbol{\lambda}^*$.

The studied trucated variable Y_T has distribution f_{YT}. Let us denote $Y \sim f_Y$ the corresponding non-truncated distribution. The trucation occurs for some real value a. This truncation may happen for some physical modelling reason. One has:

$$f_{YT}(y) = \frac{1}{1-F(a)}1_{[a,+\infty[}(y)f_Y(y).$$

The formal definition of $M_{YT}(\boldsymbol{\lambda})$ the mfg of Y_T for some $\boldsymbol{\lambda}$ is:

$$M_{YT}(\boldsymbol{\lambda}) = \frac{1}{1 - F_Y(a)} \int_a^{+\infty} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy.$$

Let us recall that we are looking for $\boldsymbol{\lambda}^*$such as:

$$\delta = \frac{M'_{YT}(\boldsymbol{\lambda}^*)}{M_{YT}(\boldsymbol{\lambda}^*)} = \frac{\int_a^{+\infty} y f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy}{\int_a^{+\infty} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy}. \tag{11}$$

When the expression does not take a practical form, one can use numerical integration to estimate the integral term. Unfortunately, for some heavy tailed distribution (for instance Gumbel distribution), this numerical integration might be complex or not possible. This is due to the multiplication by an exponential of y. The following tip helps to aviod such problems. Denoting $M_Y(\boldsymbol{\lambda})$ the mfg of the non-trucated distribution for some , one can remark that:

$$M_Y(\boldsymbol{\lambda}) = \int_{-\infty}^{+\infty} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy = \int_{-\infty}^{a} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy + \int_a^{+\infty} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy$$

Thus another expression for $M_{YT}(\boldsymbol{\lambda})$ is:

$$M_{YT}(\boldsymbol{\lambda}) = \frac{1}{1 - F_Y(a)} \left[M_Y(\boldsymbol{\lambda}) - \int_{-\infty}^{a} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy \right].$$

The integral term is much smaller in the left heavy tailed distribution case. Therefore the numerical integration (for instance using Simpson's method) is much more precise or became possible.

The same goes for $M'_{YT}(\boldsymbol{\lambda})$ which has alternative expression:

$$M'_{YT}(\boldsymbol{\lambda}) = \frac{1}{1 - F_Y(a)} \left[M'_Y(\boldsymbol{\lambda}) - \int_{-\infty}^{a} y f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy \right].$$

Finally, another form of 11 is:

$$\delta = \frac{M'_Y(\boldsymbol{\lambda}) - \int_{-\infty}^{a} y f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy}{M_Y(\boldsymbol{\lambda}) - \int_{-\infty}^{a} f_Y(y) \exp\left[\boldsymbol{\lambda} y\right] dy}. \tag{12}$$

This alternative expression may lead to more precise estimations of $\boldsymbol{\lambda}^*$ when $M_Y(\boldsymbol{\lambda})$ and $M'_Y(\boldsymbol{\lambda})$ are known (which is the case for most usual distribution) since the integral term are much smaller than in the first expression . A reference to this Appendix is made in the summary table **??**.

D Proofs of the NEF properties

In this Appendix, the details of the calculus for the Proposition 3.4 are detailed. The definition of NEF was given in the concerned Proposition.

NEF specificities : If the original density $f_i(x)$ is a NEF, then under a set of K linear constraints on $f(x)$, one has :

$$f(x) = b(x) \exp\left[x\theta - \eta(\theta)\right],$$

thus :

$$f_\delta(x) = f(x) \exp\left[\sum_{k=1}^{K} \lambda_k g_k(x) - \psi(\boldsymbol{\lambda})\right]$$

The regularization constant from (7) can be written as:

$$\psi(\boldsymbol{\lambda}) = \log \int b(x) \exp\left[x\theta + \sum_{k=1}^{K} \lambda_k g_k(x) - \eta(\theta)\right] dx \tag{13}$$

If the integral on 13 is finite, f_δ exists and is a density.

Mean twisting With a single constraint formulated as in (8), (13) becames :

$$\begin{aligned}\psi(\boldsymbol{\lambda}) &= \log \int b(x) \exp\left[x\theta + \lambda x - \eta(\theta)\right] dx \\ &= \log \int b(x) \exp\left[x\left(\theta+\lambda\right) - \eta(\theta) + \eta(\theta+\lambda) - \eta(\theta+\lambda)\right] dx\end{aligned}$$

if $\eta(\theta+\lambda)$ is well defined.

$$\begin{aligned}\psi(\boldsymbol{\lambda}) &= \left(\eta(\theta+\lambda) - \eta(\theta)\right) + \log\left[\int b(x) \exp\left[x\left(\theta+\lambda\right) - \eta(\theta+\lambda)\right]\right] dx \\ &= \eta(\theta+\lambda) - \phi(\theta)\end{aligned}$$

since

$$b(x) \exp\left[x\left(\theta+\lambda\right) - \eta(\theta+\lambda)\right] = f_{\theta+\lambda}(x)$$

with notation from (3.4), is a density of integral 1. Thus

$$\begin{aligned}f_\delta(x) &= b(x) \exp\left[x\theta - \phi(\theta)\right] \exp\left[\lambda x - \eta(\theta+\lambda) + \eta(\theta)\right] \\ &= b(x) \exp\left[x\left[\theta+\lambda\right] - \eta(\theta+\lambda)\right] = f_{\theta+\lambda}(x)\end{aligned}$$

Thus the mean twisting of a NEF of CDF $\eta(.)$ results in another NEF with mean $\eta'(\theta+\lambda) = \delta$ (constraint) and variance $\eta''(\theta+\lambda)$.

Variance twisting With a single constraint formulated as in (9), (13) thus the new distribution has density :

$$f_\delta(x) = b(x) \exp\left[x\theta + x\lambda_1 + x^2\lambda_2 - \psi(\boldsymbol{\lambda}) - \eta(\theta)\right]$$

Since $\boldsymbol{\lambda}$ is known or computed, and θ is also known, one has the variable change $z = \sqrt{\lambda_2}x$ assuming λ_2 is strict. positive (the variable change is $z = \sqrt{-\lambda_2}x$ if λ_2 is strict. neg.). Thus,

$$\begin{aligned}f_\delta(x) &= b(\frac{z}{\sqrt{\lambda_2}}) \exp\left[z^2\right] \exp\left[\frac{z}{\sqrt{\lambda_2}}\left(\theta+\lambda_1\right) - \psi(\boldsymbol{\lambda}) - \eta(\theta)\right] \\ &= \exp\left[\eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right) - \eta(\theta) - \psi(\boldsymbol{\lambda})\right] c(z) \exp\left[z\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}} - \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right)\right]\end{aligned}$$

with

$$c(z) = b(\frac{z}{\sqrt{\lambda_2}}) \exp\left[z^2\right].$$

By (7),

$$\begin{aligned}\psi(\boldsymbol{\lambda}) &= \log \int b(x) \exp\left[x\theta + x\lambda_1 + x^2\lambda_2 - \eta(\theta)\right] dx \\ &= \log \int b(\frac{z}{\sqrt{\lambda_2}}) \exp\left[z^2\right] \exp\left[\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}z - \eta(\theta) + \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right) - \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right)\right] dx \\ &= \left(\eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right) - \eta(\theta)\right) + \log \int c(z) \exp\left[\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}z - \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right)\right] dx \\ &= \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right) - \eta(\theta)\end{aligned}$$

By (14) and (14), one has :

$$f_\delta(x) = c(z) \exp\left[z\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}} - \eta\left(\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}\right)\right]$$

thus the variance twisting of a NEF results in another NEF parametrized by $\frac{(\theta+\lambda_1)}{\sqrt{\lambda_2}}$.

E Summary Table with modified distributions for mean shift

Original distribution	Modified distribution	Modified pdf $f_{i\delta}$	Link between λ^* and δ
NEF(θ)	$NEF(\theta+\lambda^*)$	$f_{i\delta}(x_i) = b(x_i)\exp\left[x_i\left[\theta+\lambda^*\right]-\eta(\theta+\lambda^*)\right]$	$\eta'(\theta+\lambda^*)=\delta$
Special case of NEF: $\mathcal{N}(\mu,\sigma)$	$\mathcal{N}(\delta,\sigma)$	$f_{i\delta}(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_i-\delta}{\sigma}\right)^2\right]$	$\lambda^* = \frac{\delta-\mu}{\sigma^2}$
Uniform distribution: $\mathcal{U}_{[a,b]}$	$\propto$ truncated exponential	$f_{i\delta}(x_i) = \frac{\lambda^*}{e^{\lambda^* b}-e^{\lambda^* a}}1_{[a,b]}(x_i)e^{\lambda x_i}$	$\delta = \frac{1}{(b-a)}\frac{e^{\lambda^* b}\left(\lambda^* b-1\right)+e^{\lambda^* a}\left(1-\lambda^* a\right)}{\lambda^*\left(e^{\lambda^* b}-e^{\lambda^* a}\right)}$
Left Tr Gaussian $\mathcal{N}_T(\mu,\sigma,a)$	$\mathcal{N}_T(\mu+\sigma^2\lambda^*,\sigma,a)$	$f_{i\delta}(x_i) = \frac{1_{[a,+\infty[}(x_i)}{1-F(a)}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu-\sigma^2\lambda^*}{\sigma}\right)^2\right]$	$\delta = \mu+\sigma^2\lambda^* - \sigma\frac{\phi\left(\frac{a-(\mu+\sigma^2\lambda^*)}{\sigma}\right)}{1-\Phi\left(\frac{a-(\mu+\sigma^2\lambda^*)}{\sigma}\right)}$
Triangle $\mathcal{T}(a,b,c)$	–	$f_{i\delta}(x_i) = \exp(x_i\lambda^*-\psi(\lambda^*))f(x_i)$	$\delta = \frac{(a-\frac{1}{\lambda^*})e^{\lambda^* a}(b-c)+(b-\frac{1}{\lambda^*})e^{\lambda^* b}(c-a)+(c-\frac{1}{\lambda^*})e^{\lambda^* c}(a-b)}{e^{\lambda^* a}(b-c)+e^{\lambda^* b}(c-a)+e^{\lambda^* c}(a-b)}$
Left Tr Gumbel $\mathcal{G}_T(\mu,\beta,a)$	–	$f_{i\delta}(x_i) = \exp(x_i\lambda^*-\psi(\lambda^*))f(x_i)$	$\delta = \frac{M'_Y(\lambda^*)-\int_{-\infty}^{a} y f_Y(y)\exp[\lambda^* y]dy}{M_Y(\lambda^*)-\int_{-\infty}^{a} f_Y(y)\exp[\lambda^* y]dy}$ with : $M_Y(\lambda^*) = \Gamma(1-\beta)\exp\left[\lambda^*\mu\right]$ $M'_Y(\lambda^*) = \Gamma(1-\beta)\exp\left[\lambda^*\mu\right]\left[\mu-\beta F^{(0)}(1-\lambda^*)\right]$

Table 7: Modified distributions for mean twisting. Note that $\Phi(.)$is the cdf of the standard normal distribution, and $\phi(.)$is its pdf.