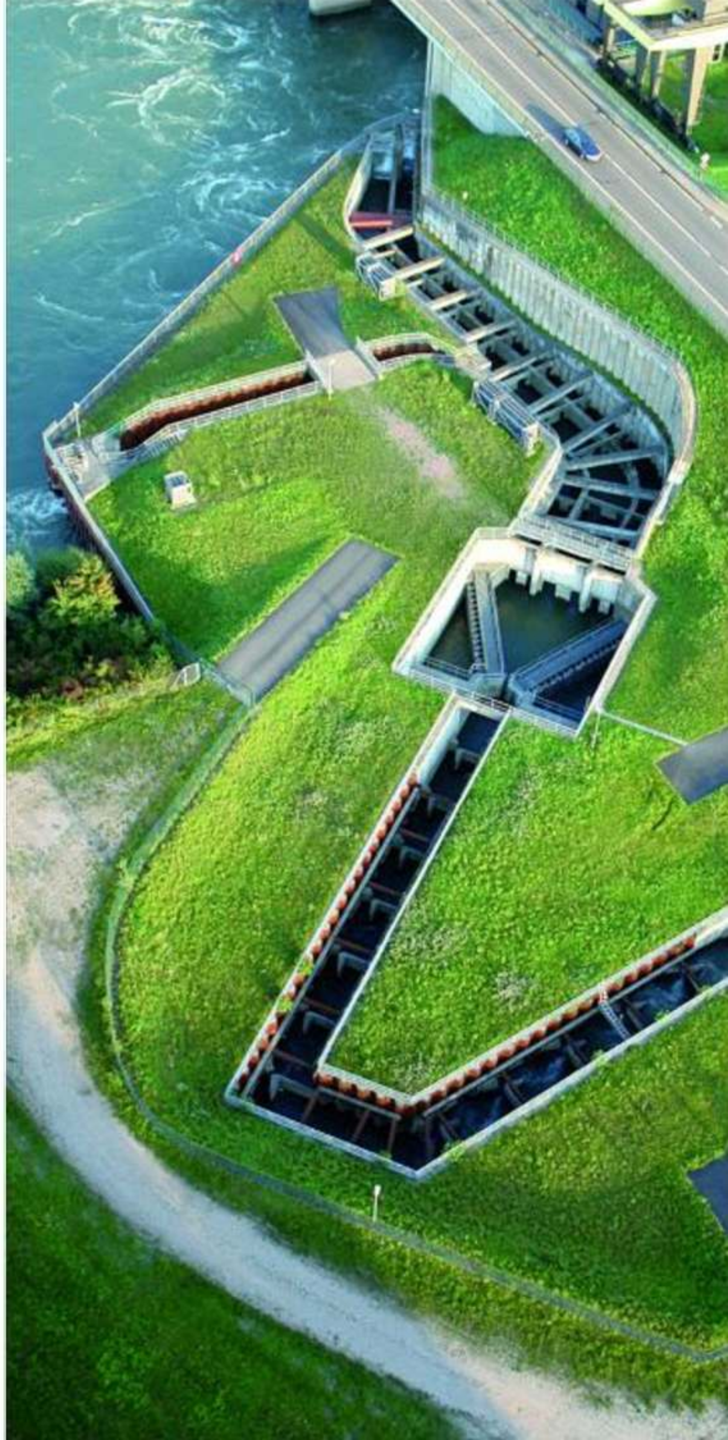
Perturbed-law based sensitivity indices: Motivations, methodology and industrial application

Bertrand IOOSS

Electricité de France (EDF) - R&D, Chatou, France

with Roman Sueur and Vanessa Vergès (EDF R&D)

ETICS 2020, Saint Pierre d'Oléron, October, 8th



Industrial motivations

Example: Simulation of IBLOCA accident

Pressurized Water Reactor scenario:
Loss of primary coolant accident due to a break in cold leg

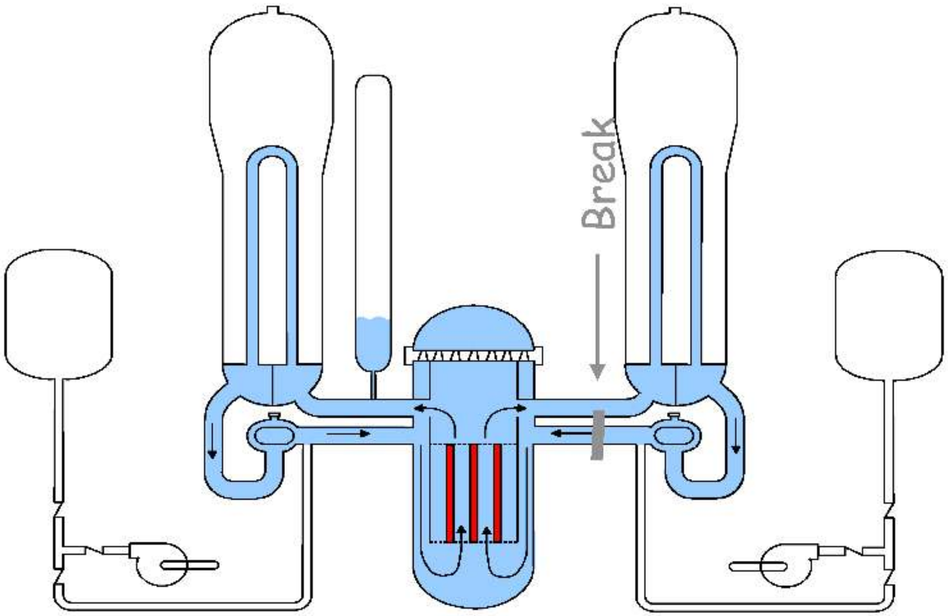
Variable of Interest :
Second peak of cladding temperature (PCT) = scalar output

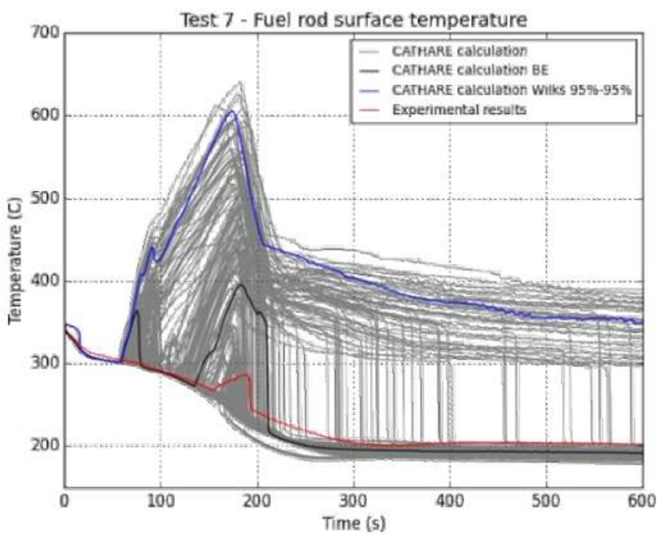
d **(~ 100) uncertain input variables :**
Critical flowrates, initial/boundary conditions, phys. eq. coef., …

<u>Modelled using</u> **<u>CATHARE code:</u>**
(thermal-hydraulic phenomena)

CPU cost for one code run > 1 hour
In industrial studies: *N* ~ O(1000) runs

QoI : **High quantiles of the PCT**





SAFETY ISSUES IN EDF ENGINEERING

Goals: Assess margins with regards to a regulatory criteria (the regulator will accept the safety approach if a sufficient margin remains)

1) Historical approach

=> **Conservative models** (e.g. without compensating physics) with **conservative inputs' values** (leading to the most penalizing calculation, corresponding to expert-based min. or max. value of each input)

2) New requirements:

Safety authorities: higher demands in terms of margin & realistic/complex physics

Operator: better control of margins (due to ageing) for better resources allocation & better maneuverability

=> **Realistic models** (at the industrial level) with **conservative inputs** => new problems due to interactions and non-monotonicity of complex physics

3) Objectives: better assessment of the real margins

=> BEPU (Best Estimate Plus Uncertainties): **realistic models & uncertain inputs**

BEPU ISSUES

BEPU approaches are well (and naturally) developed in the **probabilistic framework** (needing to define pdf of the inputs)

<u>Importance of the choice of the quantity of interest:</u>

- Probability of threshold exceedence

- High (but not extreme) quantile (95% to 99%):

  - easier to compute,

  - model computations remain in the validity domain of the computer code,

  - for the regulator, it allows to keep its fundamental safety margin (by comparison with the threshold)

BEPU ISSUES

BEPU approaches are well (and naturally) developed in the **probabilistic framework** (needing to define pdf of the inputs)

Importance of the choice of the quantity of interest:

- - Probability of threshold exceedence
- - High quantile (95% to 99%)

Key point: Presence of so-called **epistemic uncertainties**: parameters which are uncertain due to a lack of knowledge (vs. stochastic uncertainties)

The French nuclear regulatory authority ask to justify the probabilistic approach

=> Robustness of the study results towards the input distributions

REACTOR CASE STUDY: DIFFICULTIES IN A SAFETY LICENSING GOAL

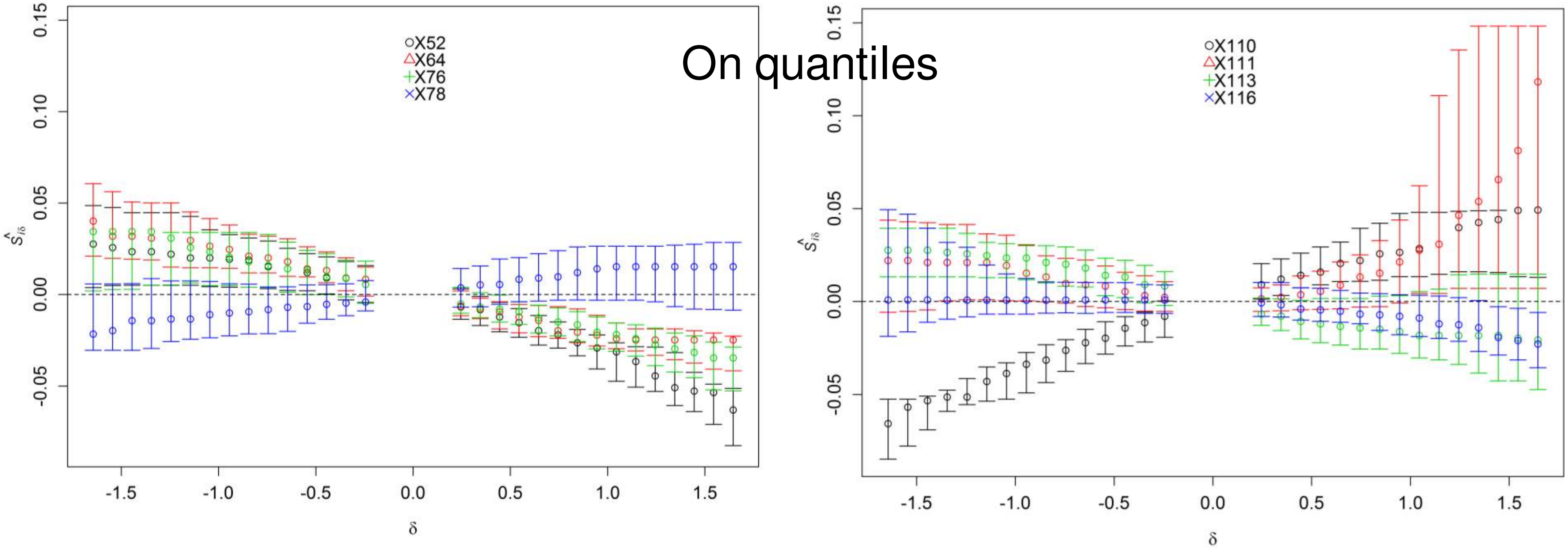
[Larget and Gautier, 2020]

Thermal-hydraulic model

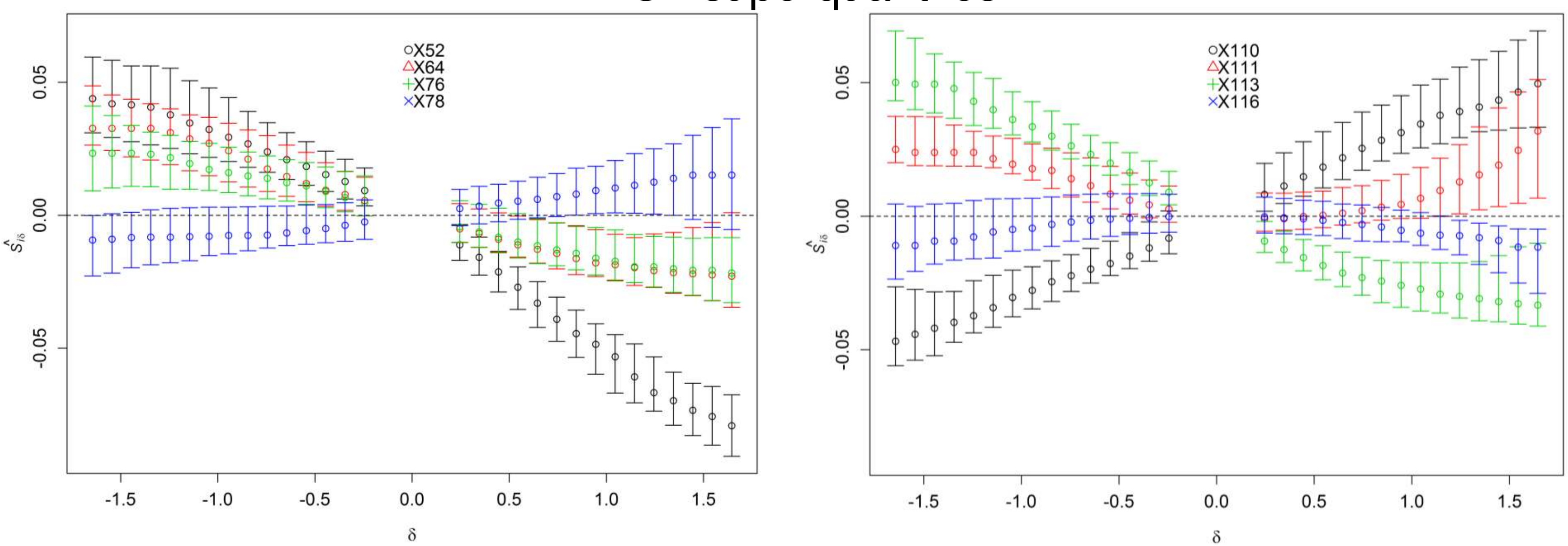
- ***d* = 100 inputs** with truncated Gaussian, log-normal, uniform, log-uniform, triangular pdf
- **2000 Monte Carlo runs**
- Quantile: q_{95} = 737°C
- Superquantile: Q_{75} = 673°C

WHICH UNCERTAIN INPUT PDF ARE INFLUENTIAL?

On quantiles



On superquantiles

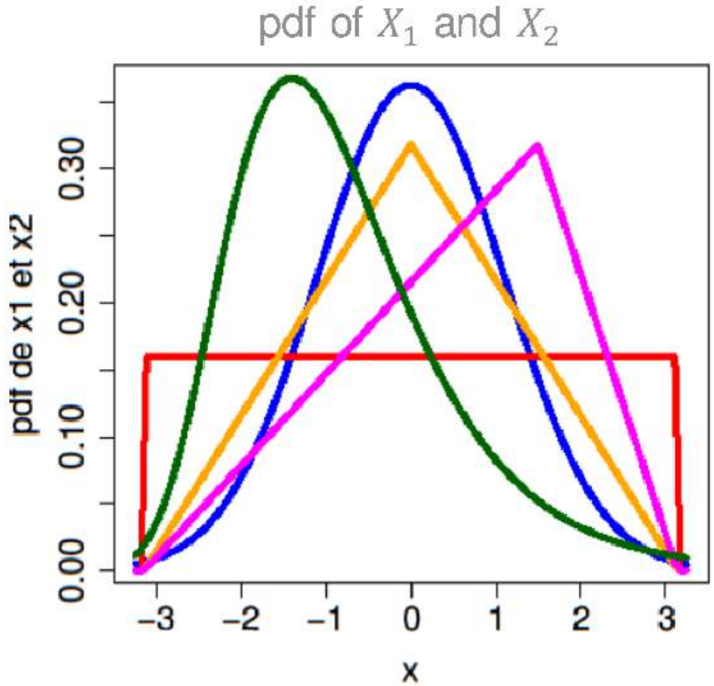


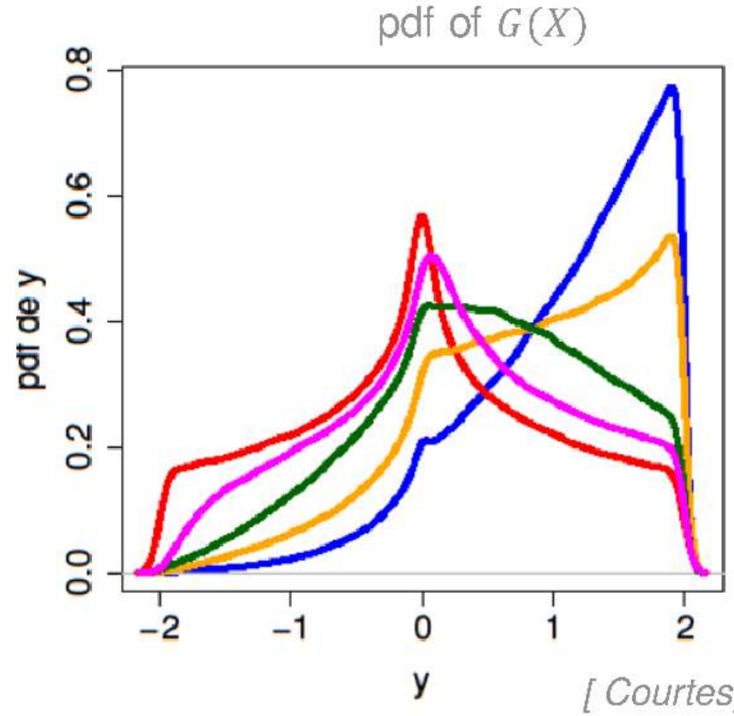
Scientific motivations

SCIENTIFIC MOTIVATION – IMPORTANCE OF INPUT PROBABILITY DISTRIBUTIONS IN UQ

$G(X) = G(X_1, X_2) = \cos(X_1) + \cos(X_2)$; X_1 and X_2 are independent with same pdf

Strong impact of the choice of the input pdf on the output distribution, and particularly on some quantities of interest: probability of exceendance, quantile, …





[Courtesy of Claire C.]

=> Needs of sensitivity analysis wrt pdf of the inputs

Global sensitivity analysis: Classical view

1. **Understand the behaviour of the model** $Y = G(X)$

2. **Simplify the computer model (dimension reduction)** **Screening**

   - **Determine the non-influential variables** (that can be fixed) and/or non-influential phenomena (to skip in the analysis)
   - Build a simplified model, a metamodel

   For example Morris, DGSM, HSIC

3. **Prioritize the uncertainty sources to reduce the model output uncertainty**

Quantitative partitioning

- Variables to be fixed to obtain the **largest output uncert. reduction**

- Most influential variables in a given output domain

For example: Sobol' indices $S_i = \frac{\mathrm{Var}(\mathrm{E}(Y|X_i))}{\mathrm{Var}(Y)}$ and $T_i = 1 - \frac{\mathrm{Var}(\mathrm{E}(Y|X_{-i}))}{\mathrm{Var}(Y)}$

Global sensitivity analysis: New view

[celui_dont_il_faut_taire_le_nom, 2020]

1. **Understand the behaviour of the model $Y = G(X)$**

2. **Simplify the computer model (dimension reduction)**

Screening

- **Determine the non-influential variables** (that can be fixed) and/or non-influential phenomena (to skip in the analysis)
- Build a simplified model, a metamodel

3. **Prioritize the uncertainty sources to reduce the model output uncertainty**

Quantitative partitioning

- Variables to be fixed to obtain the **largest output uncert. reduction**

- Most influential variables in a given output domain

4. **Analyze the robustness of the quantity of interest (QoI) with respect to the input uncertainty laws**

Robustness analysis

PERTURBED-LAW BASED INDICES (PLI)

The motivation of the PLI indices was firstly to perform global sensitivity analysis on exceedence probability computations, as classical Sobol' indices focus on contributions of input on output variance (Lemaitre, 2014)

The principle is to assess the influence of a perturbation on a parameter of the input distribution, on some quantity of interest of the model output

Recent interests:

- consideration of the quantile or superquantile as the quantity of interest
- use in industrial safety studies
- generalization of PLI

Methodology: Principles of PLI (robustness indices)

NOTATIONS

We want to study a deterministic computer code G which :

- is « costly » (CPU time, memory,…)
- has d input variables
- allows calculating the value $G(X)$ for a given set of input values $X = (X_1, \dots, X_d)$

The input variables are uncertain, hence we denote

- $\mathbb{X} \subset \mathbb{R}^d$ the domain of variation of the random vector X
- $f = \prod_{i=1}^{d} f_i$ the probability density function of X
  - each f_i is the density of X_i, the i-th marginal of X
  - the uncertain input variables $X_1, \dots X_d$ are considered independent

The QoI will be:

- a α-order quantile $q^\alpha = \inf\{t \in \mathbb{R}, F_Y(t) \geq \alpha\}$

- or a α-order superquantile $Q^\alpha = \mathbf{E}[G(X)|G(X) \geq q^\alpha]$

=> Robustness of q^α and Q^α wrt uncertainty in some f_i

PLI: THE PRINCIPLE

We aim at quantifying the impact of a perturbation on the pdf of X_i

For example, what happens if we replace $E(X_i) = \mu_i$ by $E(X_i) = \mu_i + \delta$?

We then define the PLI-quantiles as : $$S_{i\delta} = \left(\frac{q_{i\delta}^{\alpha}}{q^{\alpha}} - 1\right)$$

- It gives results in terms of percentage of perturbations
- $S_{i\delta} = 0$ when $q_{i\delta}^{\alpha} = q^{\alpha}$ i.e. when f_i has no impact on the quantile
- The sign of $S_{i\delta}$ indicates how the perturbation modifies the quantile

Remark: PLI can be defined for any QoI (e.g. PLI-superquantile and PLI for failure probability)

ESTIMATION: REVERSE IMPORTANCE SAMPLING

- We have n obs. (the model calculations) : $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)}) \rightarrow (y^{(1)}, \dots, y^{(n)})$

- **We start from classical Monte-Carlo estimators:**
  - if QoI = failure probability $(\mathrm{Proba}(Y \leq t))$: $\hat{p} = 1/N \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}}$
  - if QoI = quantile: $\hat{q}^{\alpha N} = \inf\{t \in \mathbb{R}, \hat{F}_Y^N(t) \geq \alpha\}$ where $\hat{F}_Y^N(t) = 1/N \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}}$

- Let us note $f_{i\delta}$ the perturbed density of f_i by δ, **we can estimate $\boldsymbol{p_{i\delta}}$, $\boldsymbol{F_{i\delta}(t)}$, $\boldsymbol{q_{i\delta}^{\alpha}}$ with the same sample than for $\boldsymbol{p}$, $\boldsymbol{F_Y(t)}$, $\boldsymbol{q^{\alpha}}$ by « reverse importance sampling »:**

$$\hat{p}_{i\delta}^{N} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}} L_i^{(n)} \text{ with } L_i^{(n)} = \frac{f_{i\delta}\left(x_i^{(n)}\right)}{f_i\left(x_i^{(n)}\right)}$$

[Lemaître et al., 2015]

$$\hat{F}_{i\delta}^{N}(t) = \sum_{n=1}^{N} \mathbb{I}_{\{y^{(n)} \leq t\}} L_i^{(n)} / \sum_{n=1}^{N} L_i^{(n)} \; ; \; \hat{q}_{i\delta}^{\alpha N} = \inf\{t \in \mathbb{R}, \hat{F}_{i\delta}^{N}(t) \geq \alpha\}$$

[Sueur et al., 2017]

$$\hat{Q}_{i\delta}^{\alpha N} = \frac{1}{1-\alpha} \sum_{n=1}^{N} y^{(n)} \mathbb{I}_{\{y^{(n)} \geq \hat{q}_{i\delta}^{\alpha N}\}} L_i^{(n)}$$

[I. et al., 2020]

No need of new runs of G model

PLI INDICES ESTIMATION

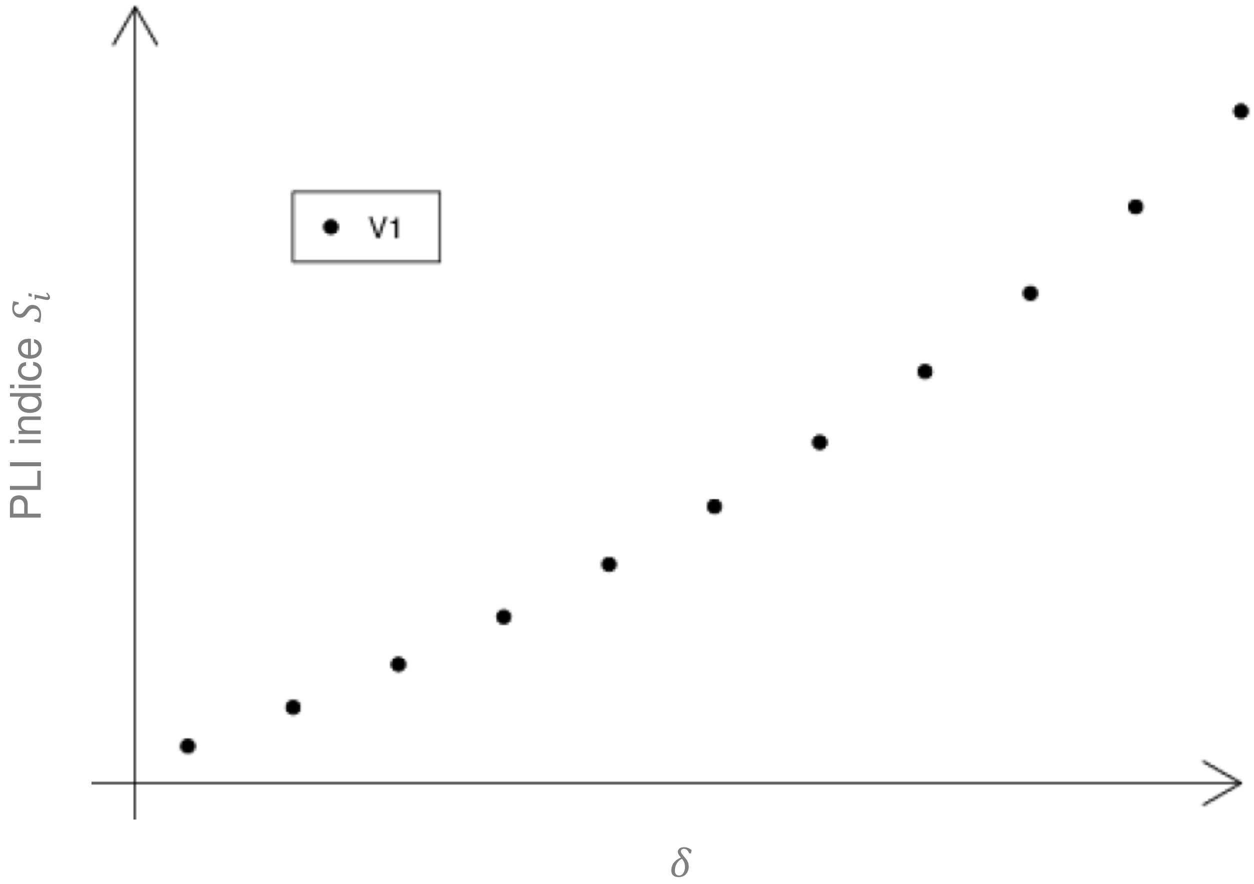
We then estimate the PLI-quantile indices with the so-called plug-in estimator:

$$\hat{S}_{i\delta}^{N} = \left(\frac{\hat{q}_{i\delta}^{\alpha N}}{\hat{q}^{\alpha N}} - 1 \right)$$

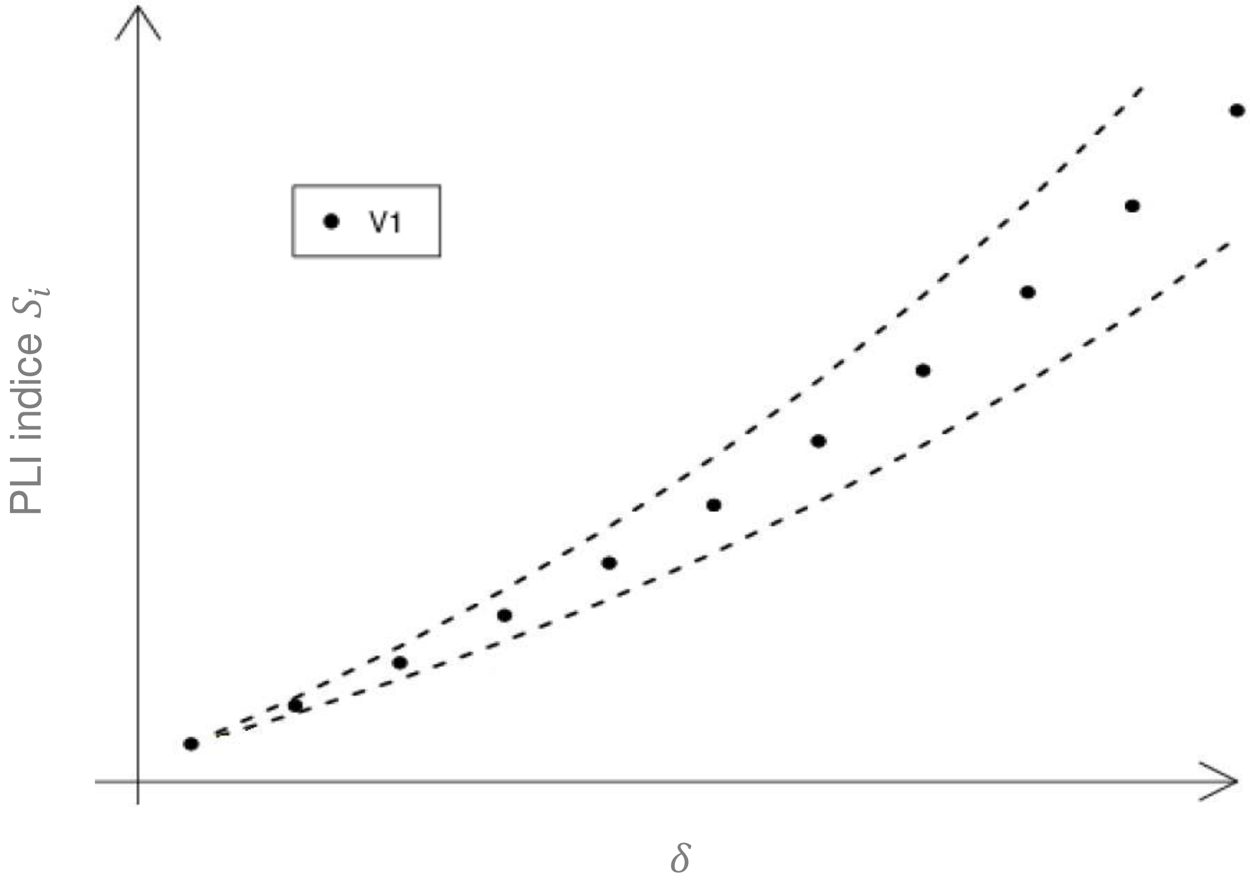
- Convergence and cental limit theorem of this estimator has been obtained (*[Lemaître et al., 2015 for failure probability; Gauchy et al., 2020 for quantile)] ,* with natural assumption $\mathrm{Supp}(f_{i\delta}) \subseteq \mathrm{Supp}(f_i)$, and others more complex:
  - For example, for PLI-quantile: $\int_{\mathrm{Supp}(f_i)} \left(\frac{f_{i\delta}(x)}{f_i(x)} \right)^3 dx < +\infty$

- For PLI-quantile, confidence intervals are easier to compute by bootstrap

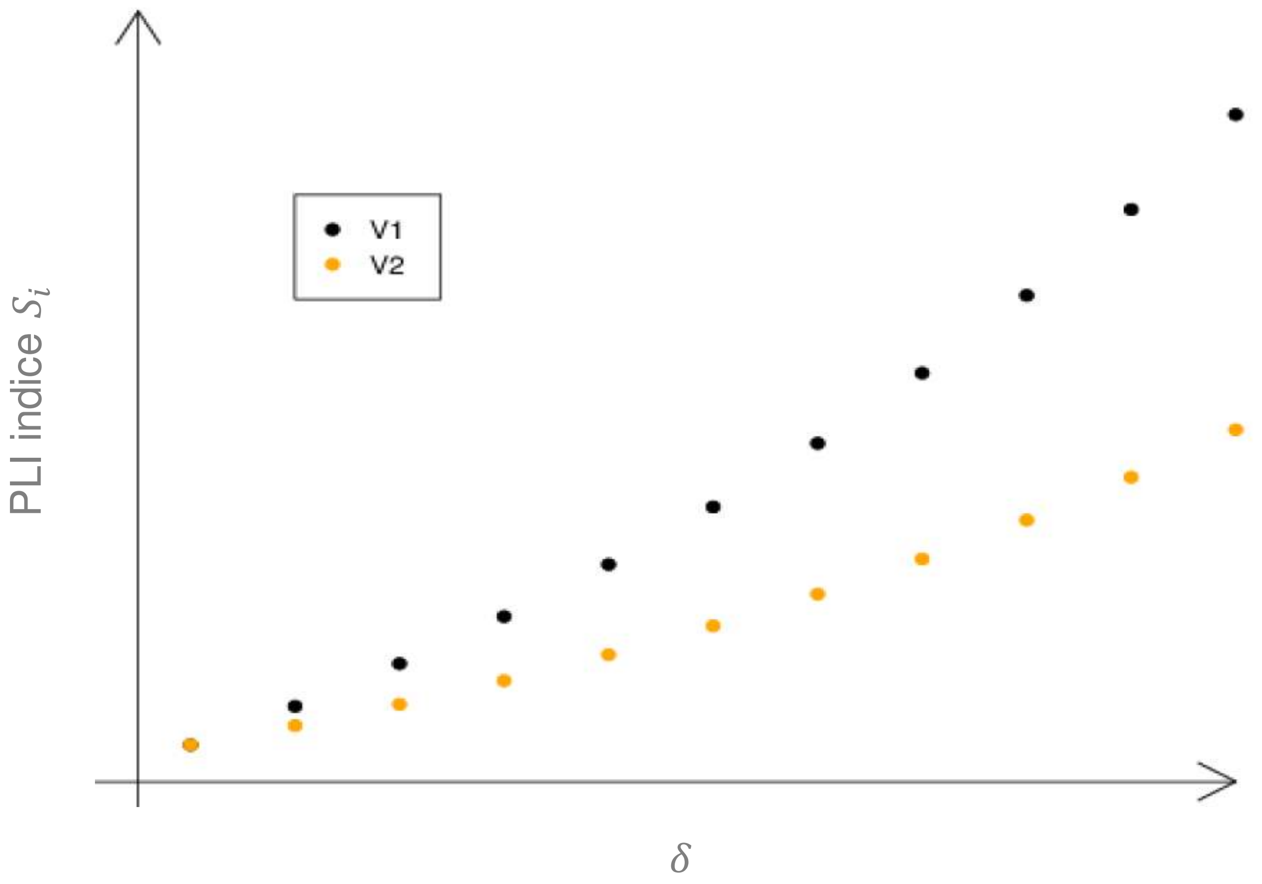
ILLUSTRATION



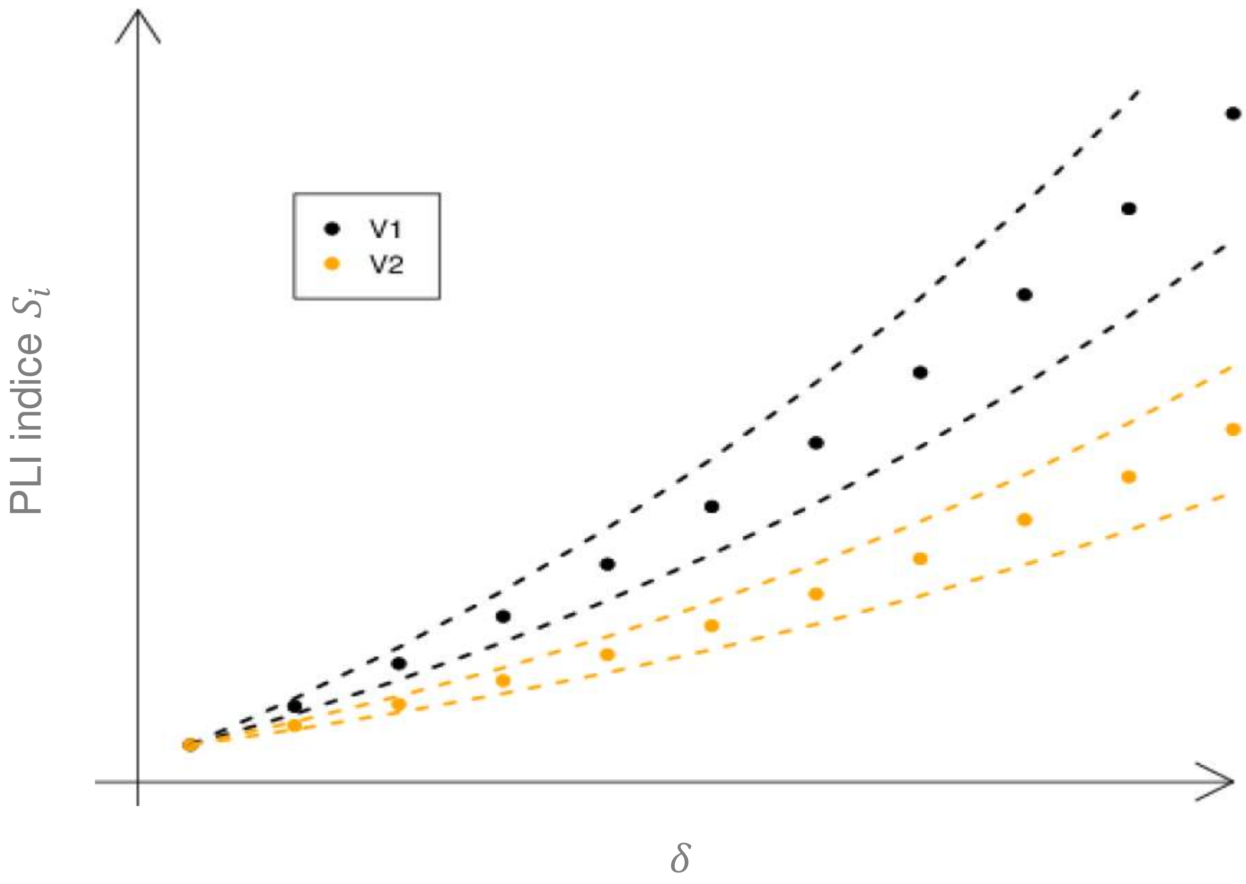
ILLUSTRATION



ILLUSTRATION

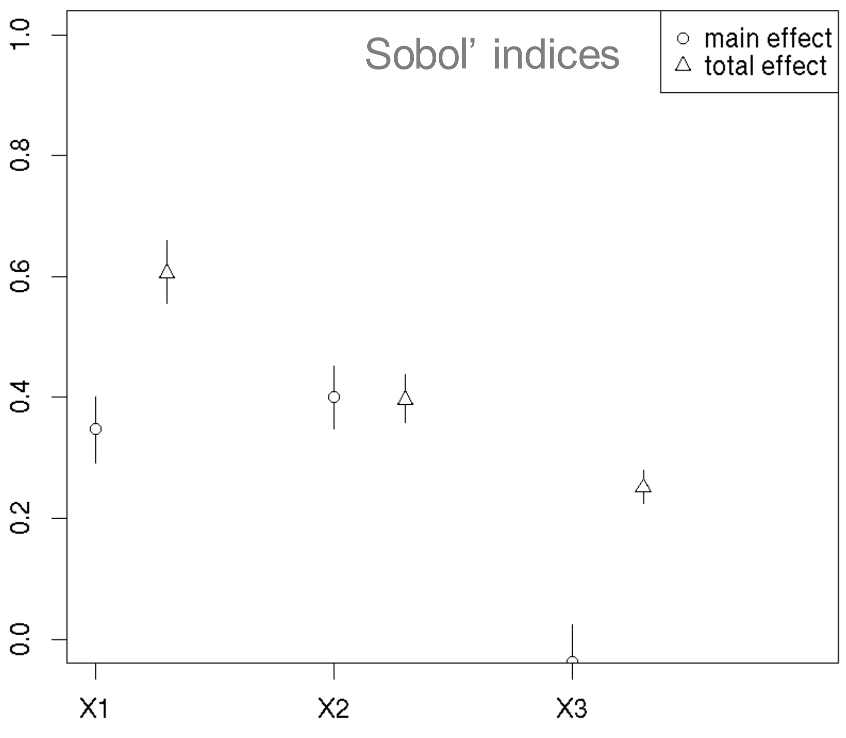


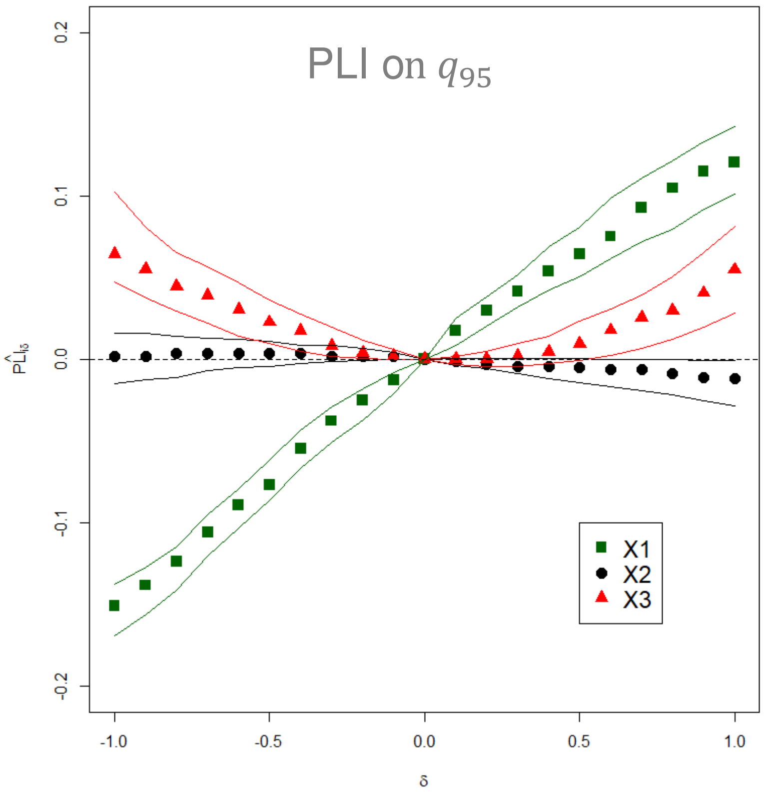
ILLUSTRATION



ANALYTICAL EXAMPLE

$$G(X) = \sin(X_1) + 7 * \sin^2(X_2) + 0,1 * X_3^4 * \sin(X_1) \; ; \quad X_i \sim U(-\pi, \pi) \text{ independent}$$





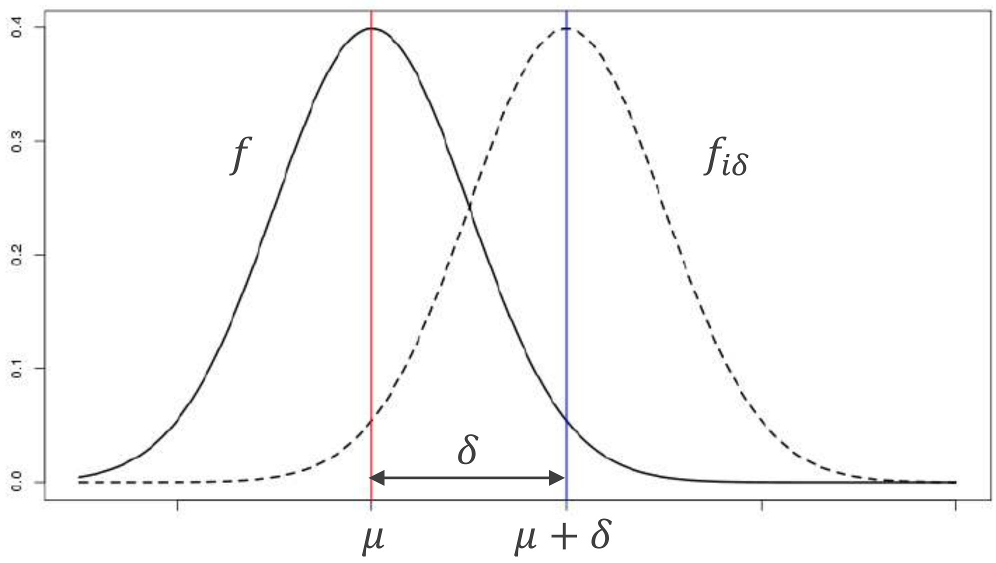
Perturbation on the pdf mean – $N = 10000$

The provided information are different

Density perturbation

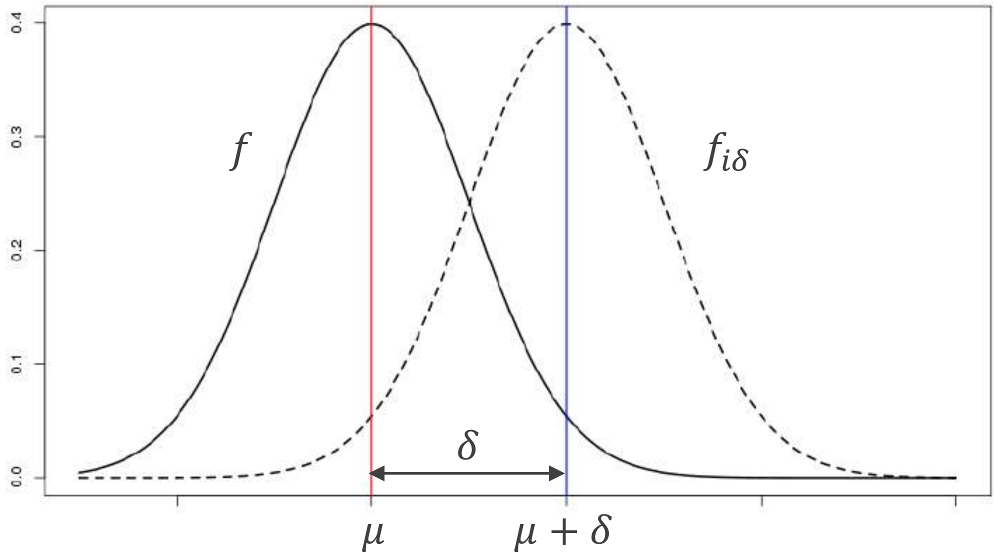
HOW TO DEFINE A DENSITY PERTURBATION ?

- Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- What if the mean of X_i was not μ but $\mu + \delta$?



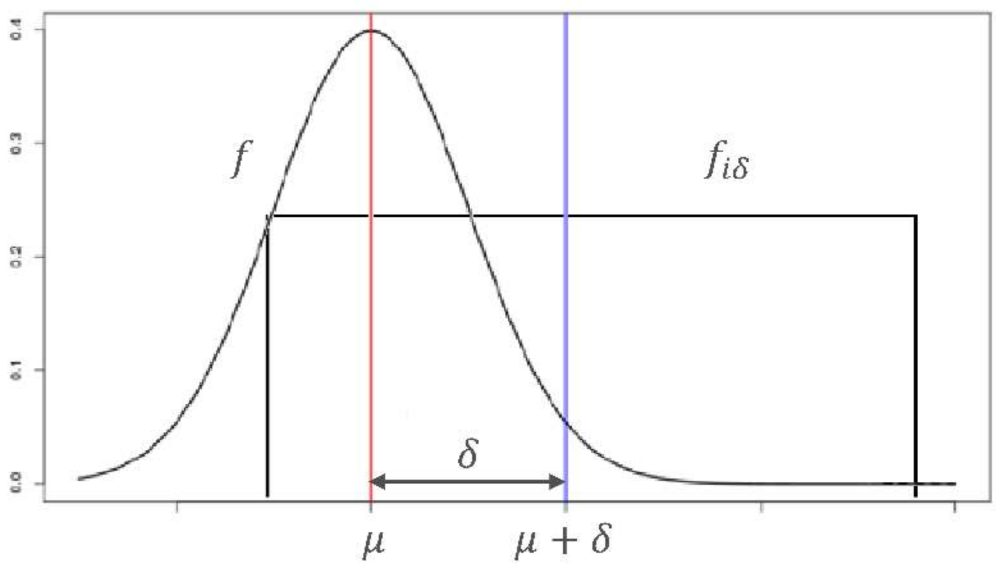
HOW TO DEFINE A DENSITY PERTURBATION ?

- **Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$**
- **How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?**



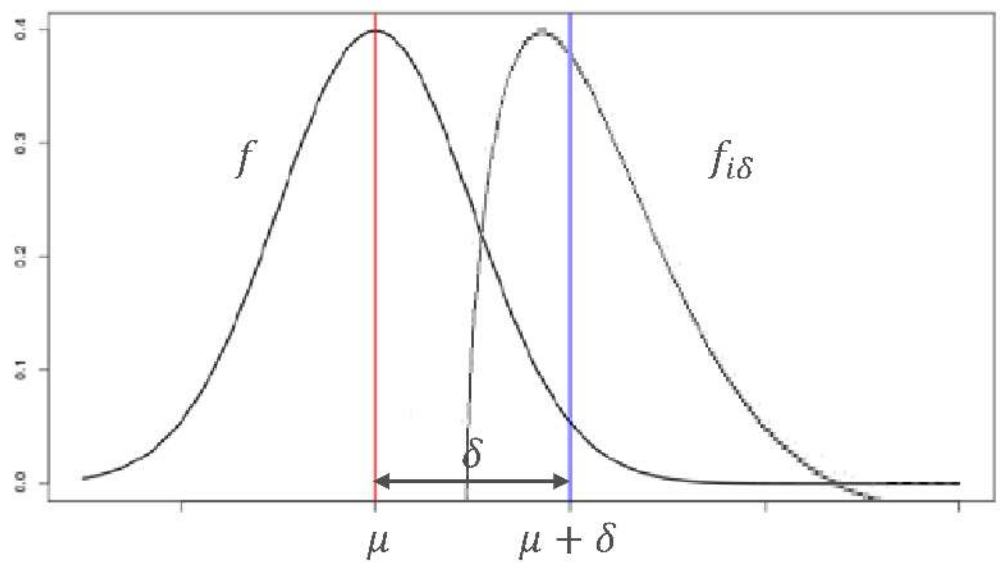
HOW TO DEFINE A DENSITY PERTURBATION ?

- Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?



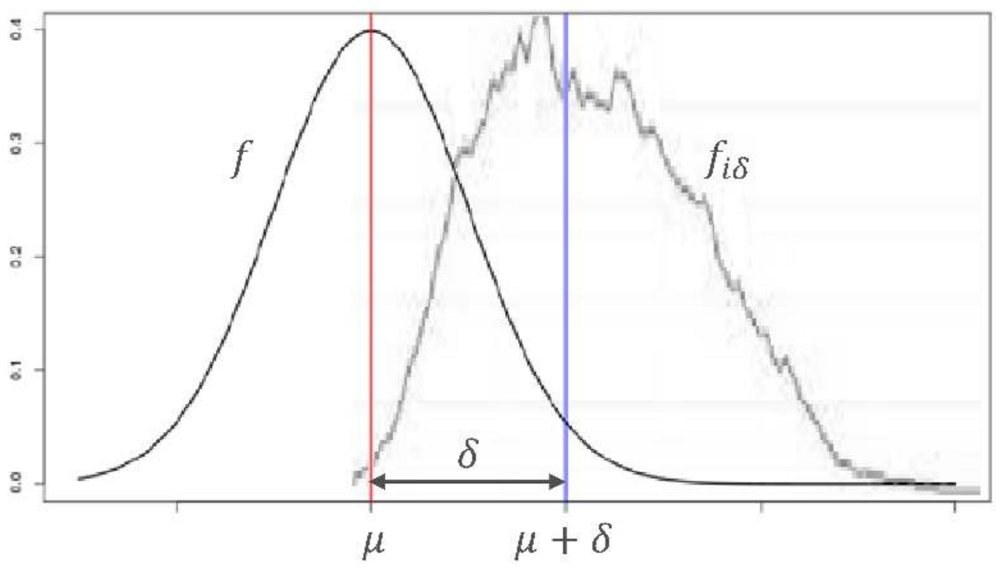
HOW TO DEFINE A DENSITY PERTURBATION ?

- Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?



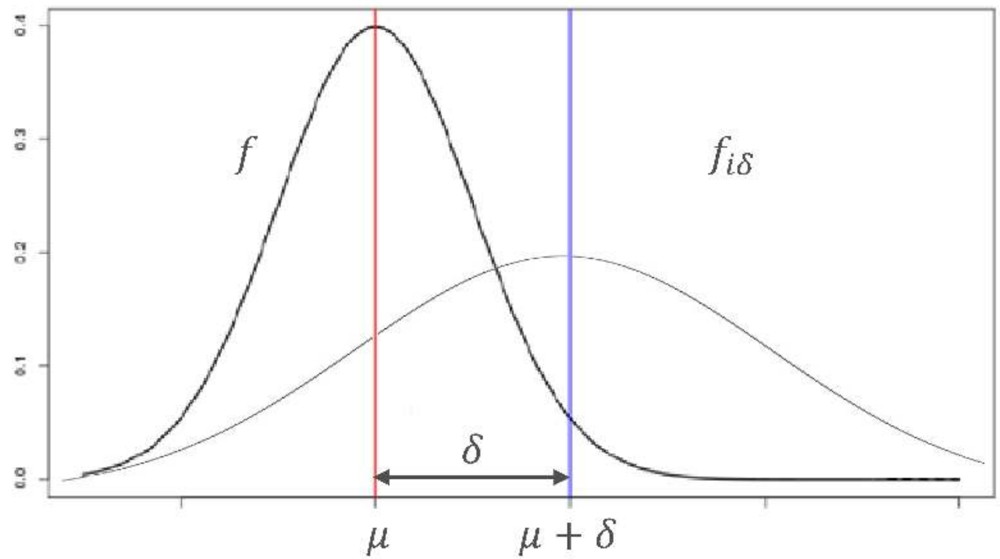
HOW TO DEFINE A DENSITY PERTURBATION ?

- Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?



HOW TO DEFINE A DENSITY PERTURBATION ?

- Let's assume that the X_i input variable has a normal distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- How to define $f_{i\delta}$ with the constraint $\int_{\mathbb{X}_i} x_i f_{i\delta}(x_i) dx_i = \mu + \delta$?



HOW TO DEFINE A DENSITY PERTURBATION ?

[Lemaître et al., 2015]

- We suggest to define the perturbed density $f_{i\delta}$ as the closest one from the initial f_i in the sense of the entropy, under the constraint of perturbation

- i.e. in the sense of Kullback-Leibler divergence :

$$KL(\pi_1,\pi_2) = \int_{-\infty}^{+\infty} \pi_1(x) \log\left(\frac{\pi_1(x)}{\pi_2(x)}\right) dx$$

- So we can give a general formal definition for $f_{i\delta}$ the following way :

$$f_{i\delta} = \underset{\substack{\pi \\ s.t.\ \mathbb{E}_\pi[g_k]=\delta_k \\ k=1,\dots,K}}{\operatorname{argmin}} KL(\pi, f_i)$$

where :

- $g_1, \dots, g_K$ are K linear constraints on the modified density
- and $\delta_1, \dots, \delta_K$ are the values for the perturbed parameters

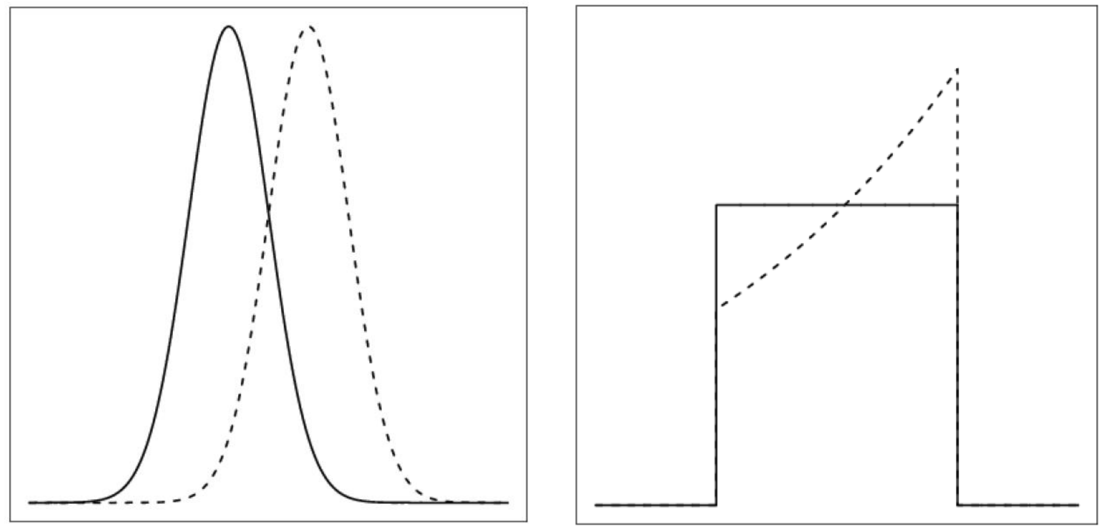
EXAMPLES OF PERTURBED PDF

Mean μ ; Variance σ^2

Gaussian | Uniform

Mean perturbation

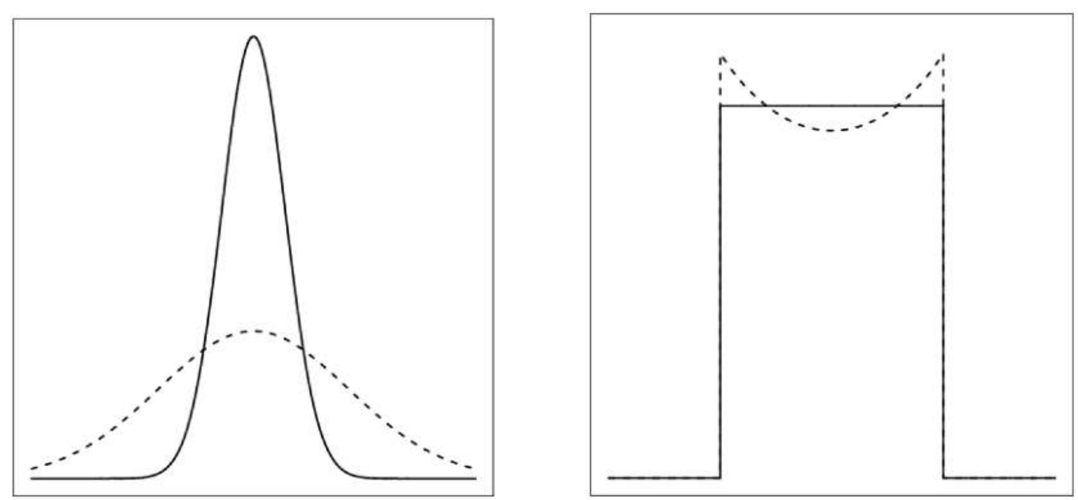
$$\mathbb{E}[X_i] = \mu + \delta$$



Variance perturbation

$$\mathbb{E}[X_i] = \mu$$

$$\mathrm{Var}[X_i] = \sigma^2 + \delta$$



This perturbation approach can be criticized [Gauchy et al., 2020]

SIMPLER PERTURBATION WAY: GENERAL DENSITY PERTURBATION

[Perrin and Defaux, 2019]

- In the case of some usual pdf, we have an analytical expression of $\frac{f_{i\delta}}{f_i}$, e.g a perturbed Gaussian pdf is another Gaussian pdf of different mean or variance
- But it is not always possible! (e.g. lognormal pdf)
- By applying an iso-probabilistic transformation (e.g. Rosenblatt transformation), we switch to the **standard space** and then get Gaussian pdf for each inputs

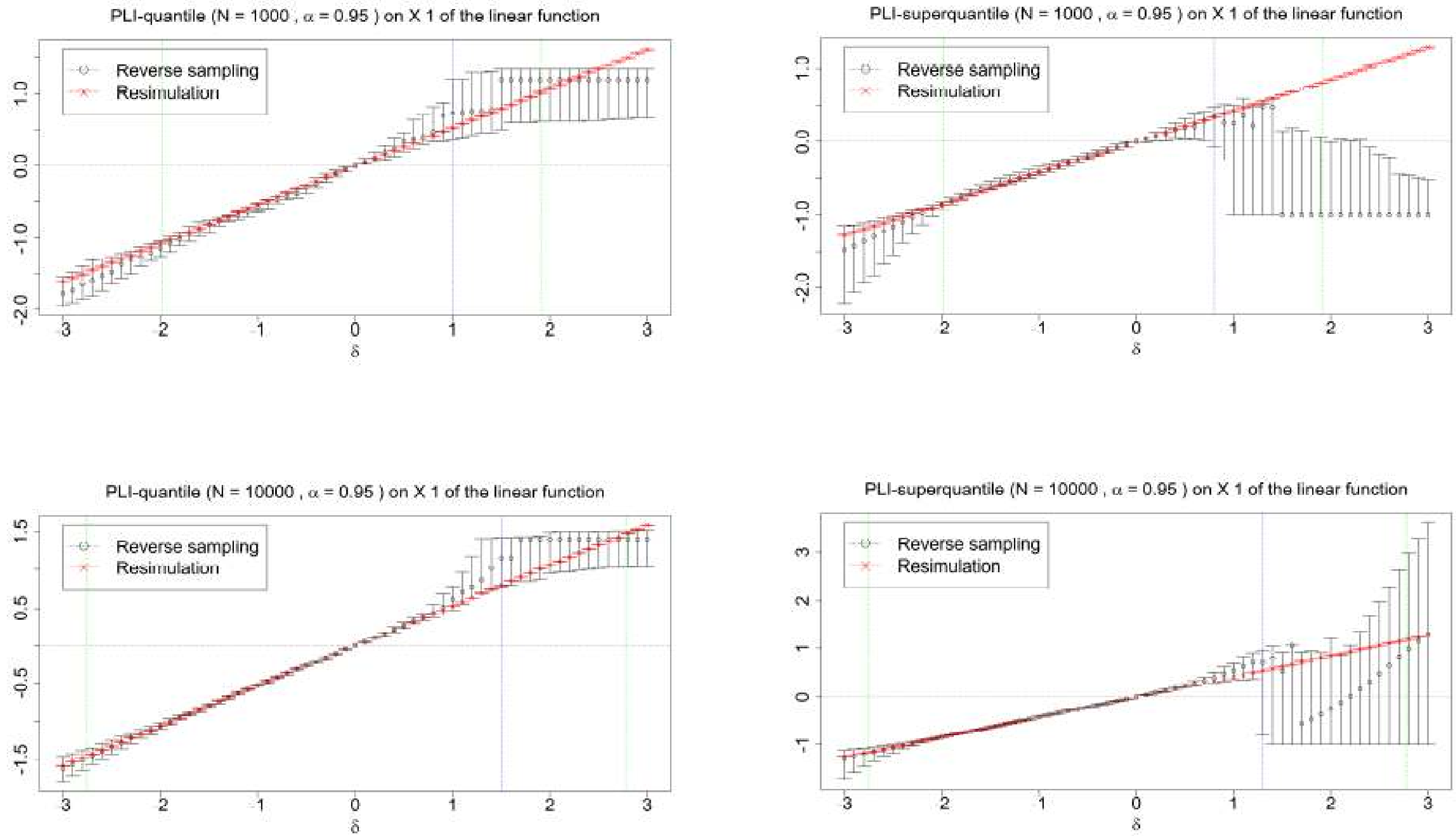
$$\Phi^{-1} \circ F_{X_i}\left(x_i^{(1)}, \dots, x_i^{(N)}\right) = \left({x'}_i^{(1)}, \dots, {x'}_i^{(N)}\right) \sim \mathcal{N}(0,1)$$

- PLI indices can then be easily determined
- However, interpreting the results in the (initial) physical space can be difficult

[Gauchy et al., 2020]

Validation of PLI-quantiles and PLI-superquantiles

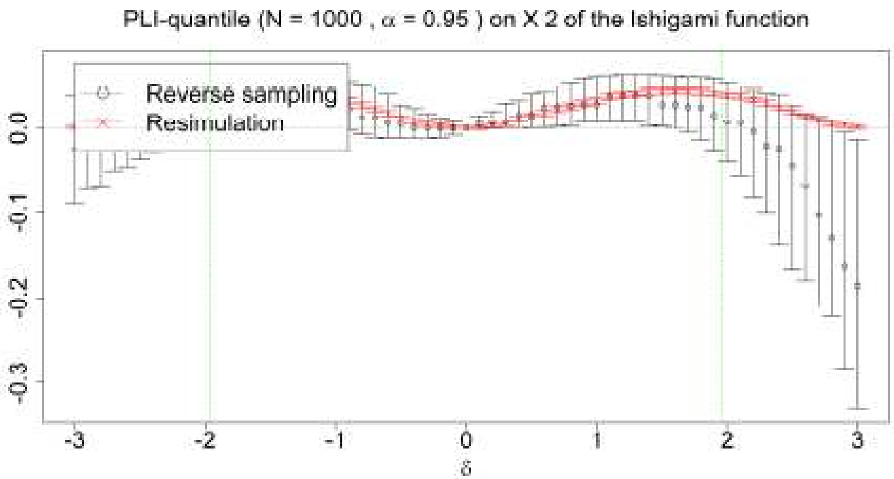
LINEAR MODEL: $Y = 2X_1 + X_2 + X_3/2,\ X_i \sim \mathcal{N}(0,1)$

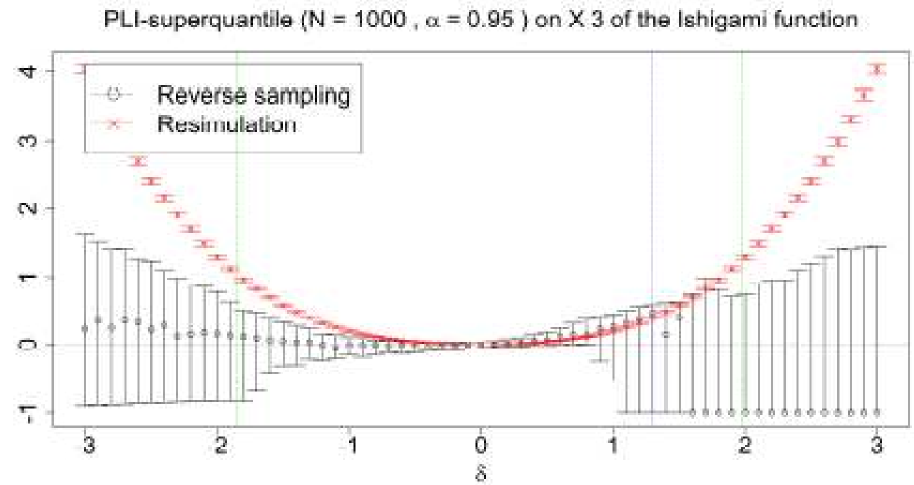


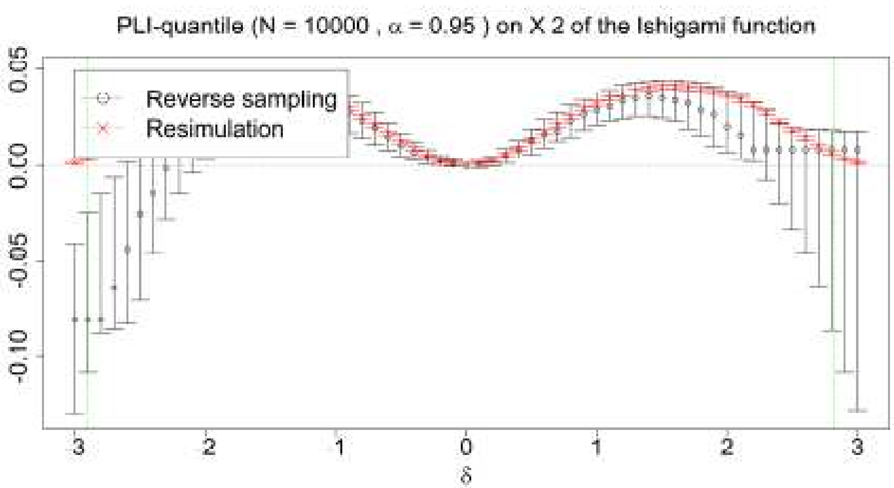
<u>Les barres vertes et bleues sont des heuristiques qui donnent des limites de domaine de validité:</u>

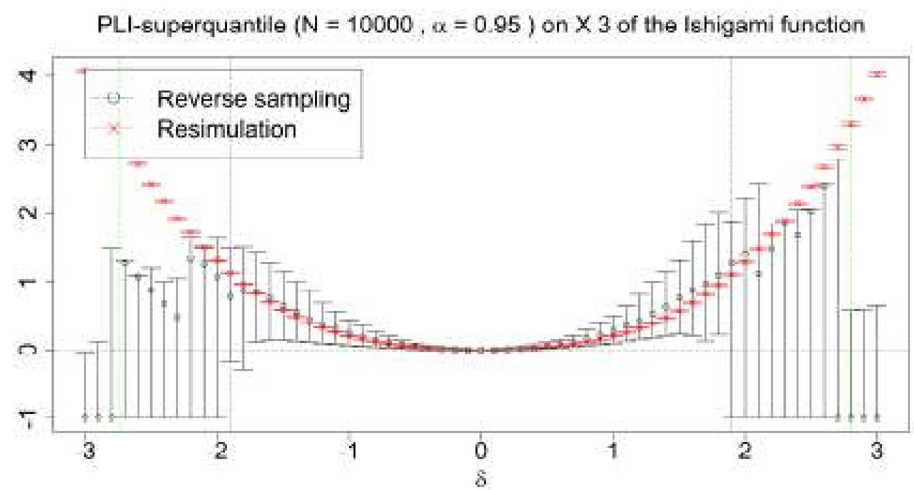
- Verte : nb de pts minimal, ici 30, à avoir de chaque côté de la moyenne de l'entrée perturbée
- Bleue : nb de pts minimal, ici 5 (resp. 10) pour le quantile (resp. superquantile) à avoir de chaque côté du quantile (resp. superquantile) perturbé

ISHIGAMI FCT: $Y = \sin(X_1) + 7\sin^2(X_2) + 0.1X_3^4 \sin(X_1),\ X_i \sim \mathcal{N}(0,1)$





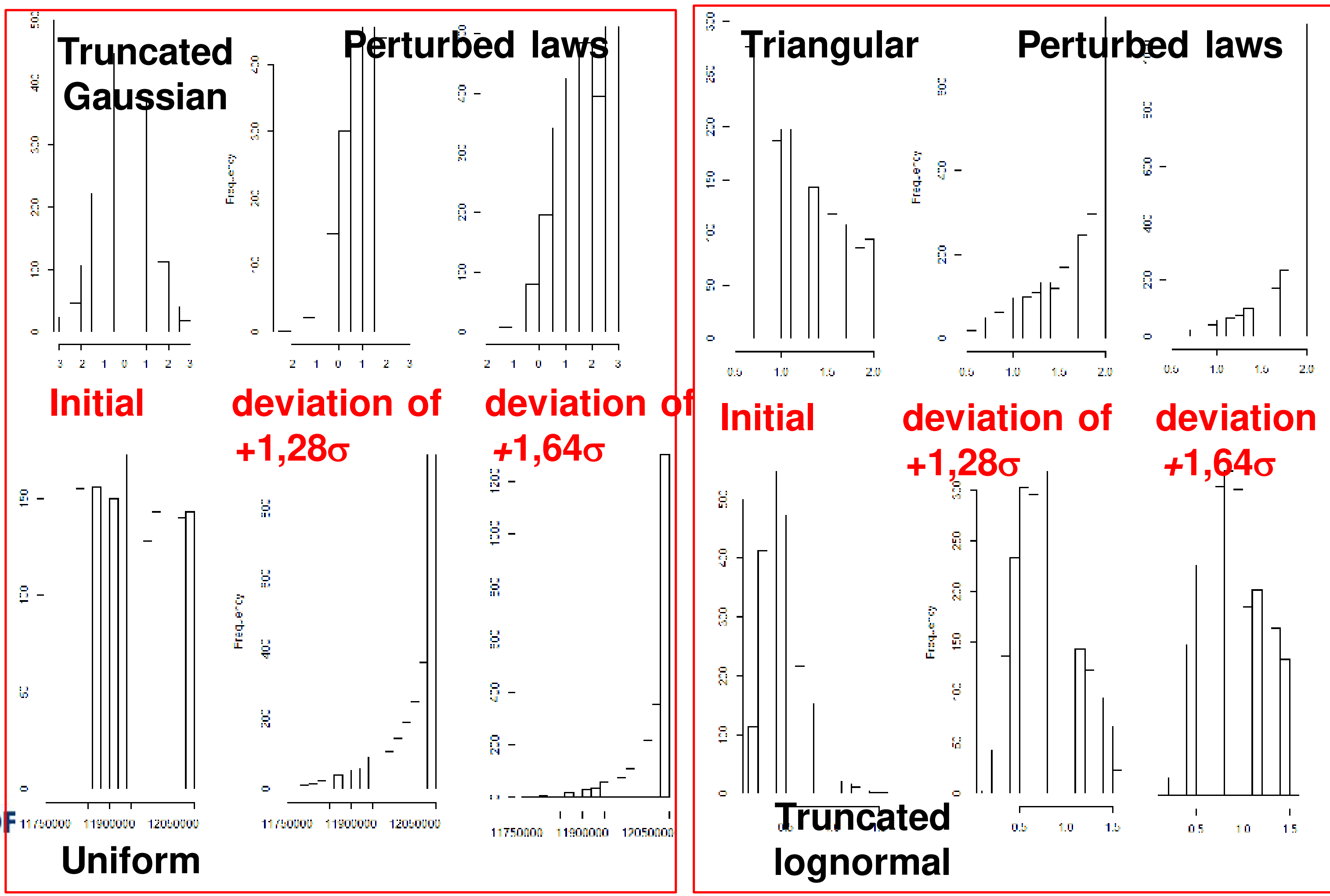




Applications

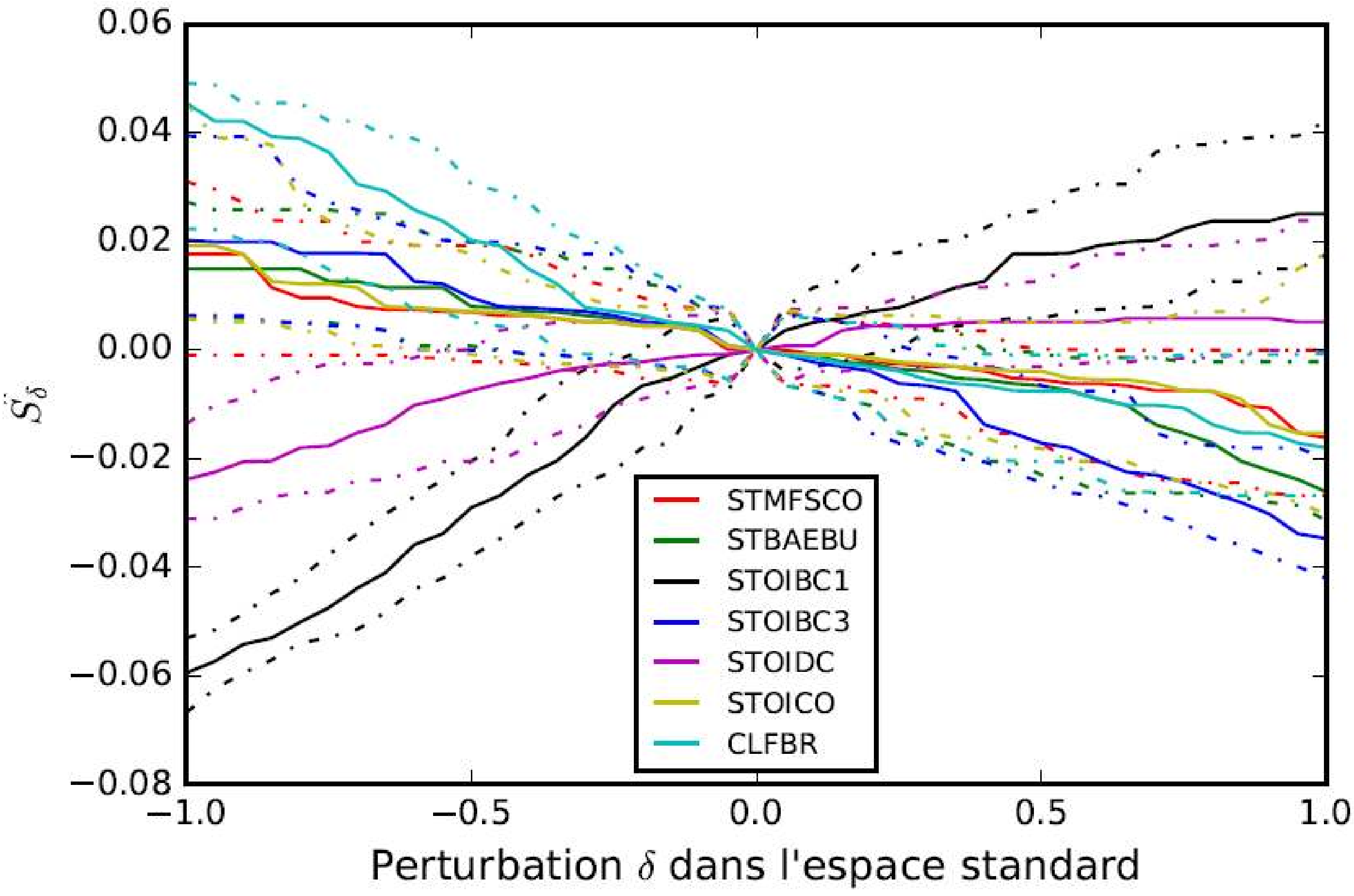
THERMAL-HYDRAULIC MODEL OF A MOCKUP

- 27 inputs with truncated Gaussian, log-normal, uniform, log-uniform, triangular pdf
- Monte-Carlo sampling of 1000 runs
- Perturbation on the mean between [-1;1] in the standard space (each input ~$\mathcal{N}(0,1)$)



RESULTS

- Graphs show the PLI of the 7 most influential variables
- 90%-confidence intervals are obtained by bootstrap



A few conclusions

- Quantile seems to be robust towards the pdf: less than 5% variation

- Sign of the PLI allows to know which value allows us to be conservative

- Non-monotonic behaviour (STOIDC)

Conclusion

CONCLUSION: BENEFITS OF PLI

- Allows to quantify the robustness of a QoI of a model output wrt uncertainty on inputs' pdf parameters (mean and variance)
- Confidence Intervals (CLT for probability of exceedance, bootstrap for quantile/superquantile) and simple heuristics allow to adjust the calculation budget (number of runs of the G code)
- No need of new runs of the G code and the input dimension is not an issue
- Easy to perturb several inputs at the same time

- Software implementation: R package 'sensitivity'

Many open issues (to be discussed)

REFERENCES

S. Da Veiga, F. Gamboa, B. Iooss and C. Prieur. *Basics and trends in sensitivity analysis: Theory and practice in R*, submitted, 2020

C. Gauchy and J. Stenger and R. Sueur and B. Iooss, An information geometry approach for robustness analysis in uncertainty quantification of computer codes, Preprint - hal-02425477

B. Iooss, V. Vergès and V. Larget, BEPU robustness analysis via perturbed-law based sensitivity indices, *BEPU 2020*, Sicily, Italy, June 2021 - hal-02864053

V. Larget and M. Gautier, Increasing conservatism in BEPU IB LOCA safety studies using complementary and industrially cost effective statistical tools, *BEPU 2020*, Sicily, Italy, June 2021

P. Lemaître. *Sensitivity analysis in structural reliability*, Thèse de l'Université de Bordeaux I, 2014

P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa and B. Iooss. Density modification based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85 :1200-1223, 2015

G. Perrin and G. Defaux, Efficient estimation of reliability-oriented sensitivity indices," *Journal of Scientic Computing*, 80, 2019

R. Sueur, B. Iooss and T. Delage. Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case, *10th International Conference on Mathematical Methods in Reliability (MMR 2017),* Grenoble, France, July 2017