Development of a Distributional Robustness Method for Analysis of Computer Codes

RT-UQ, PhD students day, Grenoble

B. B. Ketema[1,2], N. Bousquet[1,3,4], F. Costantino[2], F. Gamboa[2], B. Iooss[1,2,3], R. Sueur[1]

[1]EDF R&D, [2]IMT, [3]SINCLAIR AI Lab, [4]LPSM

April 22nd 2025

Motivation #1: a flood model

The **height** H of a river is computed using a simplified flood model

$$H = (Q/30K)^{0.6} 50^{0.3} (Z_m - Z_v)^{0.3}$$

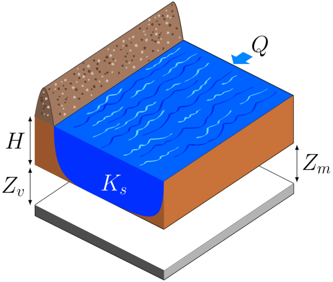
with **uncertain** inputs K, Q, Z_m, Z_v having **nominal distributions**

- $K \sim$ truncated normal
- $Q \sim$ truncated Gumbel
- $Z_m, Z_v \sim$ triangulars

Flooding does not occur 95% of the time if

$$q_{0.95}(H) \leq H_*$$

where H_* is the altitude of the dyke



Source: [Ste20]

Motivation #2: a hydrological rainfall-runoff model

The **runoff volume** R is computed using a simplified SCS hydrological model [Dav19]

$$R = (P - 0.2S)^2(P + 0.8S)^{-1}$$

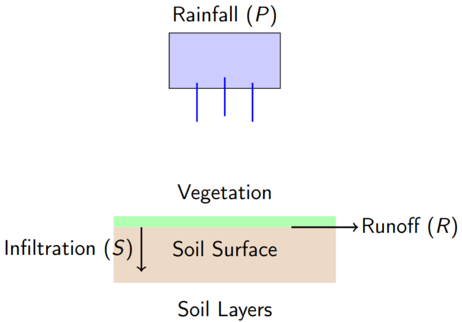
with uncertain inputs P, S having **nominal distributions**

- $P \sim$ Gumbel
- $S \sim$ truncated normal

Flooding does not occur 95% of the time if

$$q_{0.95}(R) \leq R_*$$

where R_* is the maximum drainage capacity of the urban area



Motivation #3: a thermal hydraulic computer code (nuclear context)

The **peak cladding temperature** T after a LOCA accident is computed using CATHARE code

$$T = G(X_1, \ldots, X_d)$$

where $X_1, \ldots, X_d$ are **uncertain** physical inputs having **nominal distributions**

- truncated normals
- truncated log-normals
- uniforms
- log-uniforms

Safety is guarantied $\alpha \cdot 100\%$ of the time if

$$q_\alpha(T) \leq T_*$$

where T_* is a safety threshold

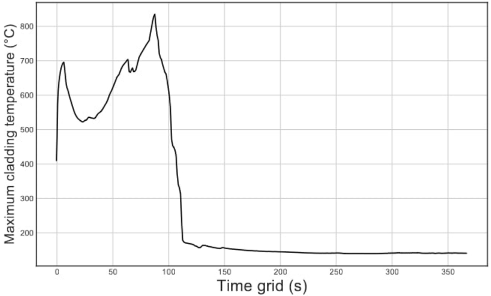


Figure: A single simulation of a transitory. *Source:* [dP21]

Commonality of previous cases

In all three cases, we have a (computer) model G

$$\boldsymbol{X} = \underbrace{(X_1, \ldots, X_d)}_{\text{uncertain input variables}} \stackrel{G}{\longmapsto} \underbrace{Y = G(\boldsymbol{X})}_{\text{output decision variable}}$$

with **nominal distribution P_0** describing uncertainty on $\boldsymbol{X}$ and a safety criterion

$$q_\alpha(Y | \boldsymbol{X} \sim \mathrm{P}_0) \leq \tau_* \tag{SC$_\alpha$}$$

Commonality of previous cases

In all three cases, we have a (computer) model G

$$\boldsymbol{X} = \underbrace{(X_1, \ldots, X_d)}_{\text{uncertain input variables}} \quad \overset{G}{\longmapsto} \quad \underbrace{Y = G(\boldsymbol{X})}_{\text{output decision variable}}$$

with **nominal distribution** P_0 describing uncertainty on $\boldsymbol{X}$ and a safety criterion

$$q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P}_0) \leq \tau_* \qquad (\mathrm{SC}_\alpha)$$

Two possible scenarios:

1. P_0 perfectly models **uncertainty** on $\boldsymbol{X}$ $\implies$ for **safety demonstrations** we only need to show (SC_α) for P_0
2. P_0 is **itself uncertain** $\implies$ important to know if (SC_α) is verified for all distributions P "neighboring" the **nominal** P_0 (think "$\mathrm{P} = \mathrm{P}_0 + \mathbf{error}$") i.e.

$$q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P}) \leq \tau_*$$

Sensitivity analysis vs. Robustness analysis

Sensitivity Analysis

Uncertainty on $\boldsymbol{X}$ is described by a **fixed** distribution P_0, the **map**

$$\boldsymbol{X} \in \mathbb{R}^d \stackrel{G}{\longmapsto} Y \in \mathbb{R} \tag{SA}$$

propagates uncertainty from $\boldsymbol{X}$ to Y, quantified using **sensitivity indices** (Sobol, HSIC,...)

Robustness Analysis

Uncertainty on P_0 itself [DVGIP21, Ste20] brings us to consider the **operator**

$$\mathrm{P} \in \mathcal{M}_1(\mathbb{R}^d) \stackrel{G_\#}{\longmapsto} G_\#\mathrm{P} \in \mathcal{M}_1(\mathbb{R}) \tag{RA}$$

Only a quantile of $G_\#\mathrm{P}$ is of interest (or another QoI). Sensitivity of $\mathrm{P} \longmapsto q_\alpha(G_\#\mathrm{P})$ around P_0 is quantified using **robustness indices** (for instance PLI, [LSA+15])

PhD project objectives

Main objective: Given G, study the sensitivity of the map

$$\mathrm{P} \in \mathcal{M}_1(\mathbb{R}^d) \longmapsto q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P})$$

around P_0, this requires to develop:

1. a **perturbation method** for changing P_0 to P (How? With what? In what space?)
2. an **estimation method** for the quantile

   $$q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P})$$

   for all distributions P "around" P_0, with a **reasonable** evaluation of G
3. a **confidence interval** $\hat{I}_{n,\beta}$ for the latter i.e.

   $$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P}) \in \hat{I}_{n,\beta}\right) \approx \beta$$

4. an **optimization algorithm** for computing robustness indices

DISTRIBUTIONAL PERTURBATION METHOD

Assumption #1: P_0 assumed **univariate** for illustration

It is not computationally feasible to verify the safety criterion (SC_α)

$$q_\alpha(Y|\boldsymbol{X} \sim \mathsf{P}) \leq \tau_*$$

for all P in $\mathcal{M}_1(\mathbb{R})$

Assumption #2: P_0 is in a **parametric family** $\mathcal{P}$ and P are restricted to it i.e.

$$\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_\theta\}_{\theta \in \Theta} \quad \text{and} \quad \mathsf{P}_0 = \mathsf{P}_{\theta_0}$$

This allows to restrict the following map to $\mathcal{P} \subset \mathcal{M}_1(\mathbb{R})$

$$\mathsf{P}_\theta \longmapsto q_\alpha(Y|\boldsymbol{X} \sim \mathsf{P}_\theta)$$

Which distance for defining perturbations ?

Definition (δ-perturbations)

Given a distance d on $\mathcal{P}$, a δ-**perturbation** of P_{θ_0} is any other P_θ in $\mathcal{P}$ such that

$$d(\mathrm{P}_{\theta_0}, \mathrm{P}_\theta) = \delta$$

In other words, all distributions P_θ in the sphere $\mathcal{S}(\mathrm{P}_{\theta_0}, \delta)$ are **equivalent** perturbations of P_{θ_0}

Most distances (or divergences) on the family $\mathcal{P} = \{\mathrm{P}_\theta\}_{\theta\in\Theta}$ are

- **extrinsic** to $\mathcal{P}$ (Wasserstein [IIBG$^+$24], total variation, f-divergences [LSA$^+$15], MMD,...)
- or **parametrization-dependent** ($\mathbb{L}^p$ distances on Θ, Mahalanobis distance,...)

A good candidate for d is the **Fisher-Rao distance** [GSSI22] which is both **intrinsic** to $\mathcal{P}$ and **invariant** under reparametrization

Definition of the Fisher-Rao distance

The **Fisher-Rao distance** has both a **statistical** and **geometric** origin

Statistics: CLT for the maximum likelihood estimator $\widehat{\theta}_n$ implies

$$n\ {}^t(\widehat{\theta}_n - \theta_*) I_{\theta_*} (\widehat{\theta}_n - \theta_*) \xrightarrow{\text{dist.}} \chi^2$$

where I_{θ_*} is the **Fisher information** for $\mathcal{P}$ defined as the Hessian of the relative entropy of $\mathbf{P}_\theta$ from $\mathbf{P}_{\theta_*}$ i.e.

$$I_{\theta_*} = \frac{\partial^2}{\partial \theta^2} \left[\mathcal{D}(\mathbf{P}_\theta | \mathbf{P}_{\theta_*}) \right]_{\theta = \theta_*}$$

Definition of the Fisher-Rao distance

The **Fisher-Rao distance** has both a **statistical** and **geometric** origin

Statistics: CLT for the maximum likelihood estimator $\widehat{\theta}_n$ implies

$$n\, {}^t(\widehat{\theta}_n - \theta_*) I_{\theta_*} (\widehat{\theta}_n - \theta_*) \xrightarrow{\text{dist.}} \chi^2$$

where I_{θ_*} is the **Fisher information** for $\mathcal{P}$ defined as the Hessian of the relative entropy of $\mathbf{P}_\theta$ from $\mathbf{P}_{\theta_*}$ i.e.

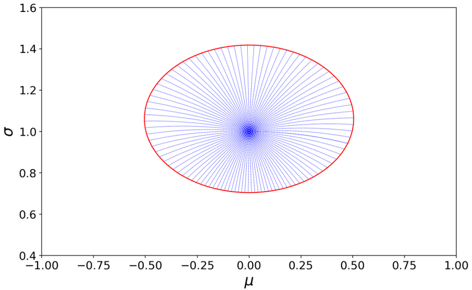
$$I_{\theta_*} = \frac{\partial^2}{\partial \theta^2} \left[\mathcal{D}(\mathbf{P}_\theta | \mathbf{P}_{\theta_*}) \right]_{\theta = \theta_*}$$

Geometry: The function $\theta \mapsto I_\theta$ can be seen as a local scalar product (Riemannian metric) on $\mathcal{P}$ allowing to define a (information) geometry. In particular, the notion of distance on $\mathcal{P}$ as the "**length of the shortest path**"

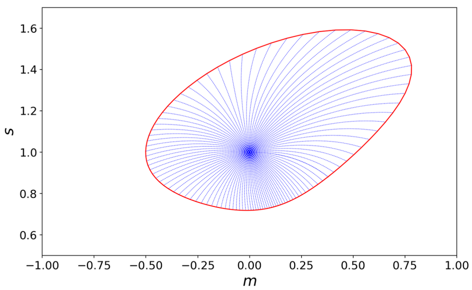
$$d(\mathbf{P}_{\theta_0}, \mathbf{P}_{\theta_1}) = \inf \left\{ \int_0^1 \sqrt{{}^t\dot{\theta}_s \cdot I_{\theta_s} \cdot \dot{\theta}_s}\, ds \;\middle|\; (\theta_s)_s \text{ curve linking } \theta_0 \text{ to } \theta_1 \right\}$$

Illustration of this perturbation method

Spheres in the **normal** and **Gumbel** family represented in the parameter space



Normal family: sphere centered at $\mathcal{N}(0,1)$ with radius $\delta = 0.5$ (μ = mean, σ = std)

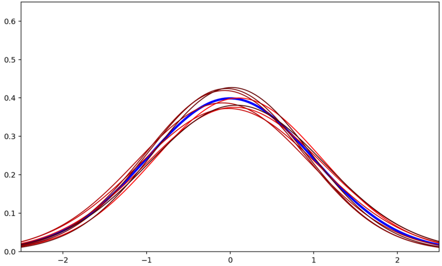


Gumbel family: sphere centered at Gumb$(0,1)$ with radius $\delta = 0.5$ (m = location, s = scale)

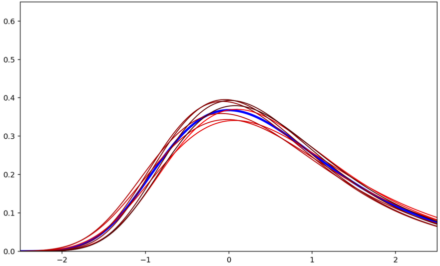
Figure: Parameters on the red sphere correspond to the **perturbed distributions P_θ** of the **nominal P_{θ_0}**

Illustration of this perturbation method

Illustration of the δ-**perturbed densities** in the **normal** and **Gumbel** family for $\delta = 0.1$



Normal family: a few $\delta-$perturbations of $\mathcal{N}(0,1)$ for $\delta = 0.1$

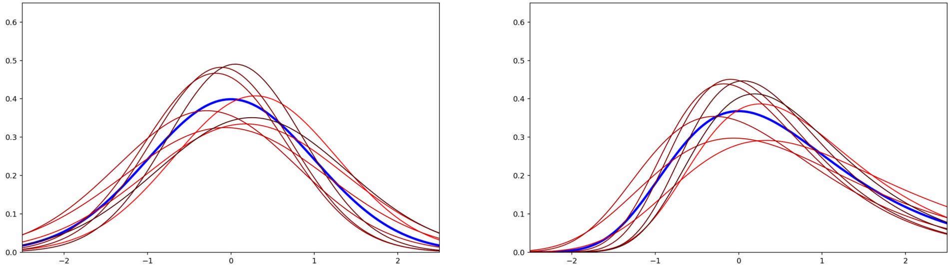


Gumbel family: a few δ-perturbations of Gumb$(0,1)$ for $\delta = 0.1$

Figure: Densities in red represent **perturbed distributions** $\mathbf{P}_\theta$ of the **nominal** $\mathbf{P}_{\theta_0}$

Illustration of this perturbation method

Illustration of the δ-**perturbed densities** in the **normal** and **Gumbel** family for $\delta = 0.3$

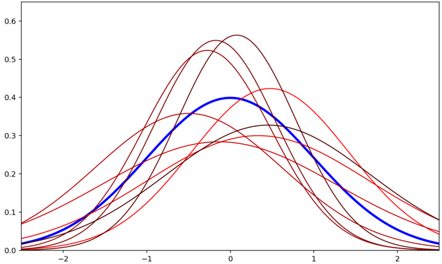


Normal family: a few $\delta-$perturbations of $\mathcal{N}(0,1)$ for $\delta = 0.3$

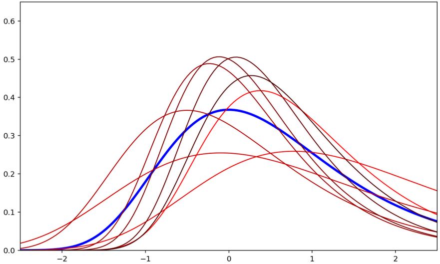
Gumbel family: a few δ-perturbations of Gumb$(0,1)$ for $\delta = 0.3$

Figure: Densities in red represent **perturbed distributions $\mathbf{P}_\theta$** of the **nominal $\mathbf{P}_{\theta_0}$**

Illustration of the δ-**perturbed densities** in the **normal** and **Gumbel** family for $\delta = 0.5$



Normal family: a few $\delta-$perturbations of $\mathcal{N}(0,1)$ for $\delta = 0.5$



Gumbel family: a few δ-perturbations of Gumb$(0,1)$ for $\delta = 0.5$

Figure: Densities in red represent **perturbed distributions** $\mathbf{P}_\theta$ of the **nominal** $\mathbf{P}_{\theta_0}$

ESTIMATION AND CONFIDENCE INTERVALS FOR QUANTILES

Direct vs. Importance sampling estimation

Reminder: Study the sensitivity of the map

$$\mathrm{P} \longmapsto q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P})$$

for perturbed distributions P around P_0

Direct method	Importance sampling
1. sample $\mathcal{X} = \{\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n\}$ from P 2. evaluate G on $\mathcal{X}$ 3. estimate $q_\alpha(Y\|\boldsymbol{X} \sim \mathrm{P})$ with the empirical quantile	1. sample $\mathcal{X}_0 = \{\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n\}$ from P_0 2. evaluate G on $\mathcal{X}_0$ 3. estimate $q_\alpha(Y\|\boldsymbol{X} \sim \mathrm{P})$ with an importance sampling technique
Not feasible for all P if G is expensive	P_0 is used as an instrumental measure in an importance sampling scheme

Quantile estimation

Henceforth, we assume $\mathbf{P} \ll \mathbf{P}_0$ and denote the likelihood ratio

$$L(x) := \frac{d\mathbf{P}}{d\mathbf{P}_0}(x)$$

of $\mathbf{P}$ against $\mathbf{P}_0$. The quantile $q_\alpha(Y|\boldsymbol{X} \sim \mathbf{P})$ is estimated as

$$\widehat{q}_\alpha(\mathbf{P}) = \inf\{t \in \mathbb{R} \mid \widehat{F}(t) \geq \alpha\}$$

where $\widehat{F}$ is the self-normalized estimator of the **distribution function** F of $Y|\boldsymbol{X} \sim \mathbf{P}$

$$\widehat{F}(t) = \frac{1}{\sum_{i=1}^{n} L(\boldsymbol{X}^i)} \sum_{i=1}^{n} L(\boldsymbol{X}^i) \mathbf{1}_{G(\boldsymbol{X}^i) \leq t}$$

which is built on the fixed sample $\mathcal{X}_0 = \{\boldsymbol{X}^1, \dots, \boldsymbol{X}^n\}$ from $\mathbf{P}_0$

Central limit theorem and confidence intervals

Theorem (CLT for IS quantile estimator, [Gly96])

If F is differentiable with positive derivative at $q_\alpha(Y|\boldsymbol{X} \sim P)$, then

$$\sqrt{n}\left(\widehat{q}_\alpha(P) - q_\alpha(Y|\boldsymbol{X} \sim P)\right) \xrightarrow{dist.} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{\mathbb{E}_{P_0}[L(\boldsymbol{X})^2(\mathbb{1}_{G(\boldsymbol{X}) \leq q_\alpha(Y|\boldsymbol{X} \sim P)} - \alpha)^2]}{F'(q_\alpha(Y|\boldsymbol{X} \sim P))}$$

Denoting $\hat{I}_{n,\varepsilon} = \widehat{q}_\alpha(\mathrm{P}) \pm \varepsilon/\sqrt{n}$, the CLT implies an **asymptotic confidence interval**

$$\lim_{n \to \infty} \mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim \mathrm{P}) \in \hat{I}_{n,\varepsilon}\right) = (2\pi\sigma^2)^{-1/2} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{x^2}{2\sigma^2}} dx$$

A consistent estimator of σ^2 is given in [CN10]

A generalization to Wilks approach

If $Y_1, \ldots, Y_n \sim G_{\#}\mathbf{P}_0$, then $q_\alpha(Y|\boldsymbol{X} \sim \mathbf{P}_0) \approx Y_{([n\alpha])}$ and

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim \mathbf{P}_0) \in \left[Y_{(i)}, Y_{(j)}\right]\right) = I_\alpha(i, n-i+1) - I_\alpha(j, n-j+1)$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ and $I_\alpha(p, q)$ is the incomplete Beta function [DN04]

A generalization to Wilks approach

If $Y_1, \dots, Y_n \sim G_\# \mathbf{P}_0$, then $q_\alpha(Y|\boldsymbol{X} \sim \mathbf{P}_0) \approx Y_{([n\alpha])}$ and

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim \mathbf{P}_0) \in \left[Y_{(i)}, Y_{(j)}\right]\right) = I_\alpha(i, n-i+1) - I_\alpha(j, n-j+1)$$

where $Y_{(1)} \leq \dots \leq Y_{(n)}$ and $I_\alpha(p, q)$ is the incomplete Beta function [DN04]

Theorem (Non-asymptotic CI for perturbed quantiles)

For n fixed and $Y_1, \dots, Y_n \sim G_\# P_0$ we have

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim P) \in \left[Y_{(i_\varepsilon)}, Y_{(j_\varepsilon)}\right]\right) \geq \beta_{n,\varepsilon}$$

where

- $\beta_{n,\varepsilon}$ *depends on assumptions on the likelihood* $\boldsymbol{L}$
- $Y_{(i_\varepsilon)} = \widehat{q}_{\alpha-\varepsilon}(P)$ *and* $Y_{(j_\varepsilon)} = \widehat{q}_{\alpha+\varepsilon}(P)$

Consequence of Hoeffding's concentration inequality

For instance, a **bounded likelihood** gives the following

Proposition

If $a \leq \boldsymbol{L} \leq b$ we have

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim \boldsymbol{P}) \in \hat{I}_{n,\varepsilon}\right) \geq 1 - \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2(\alpha-\varepsilon)^2}\right) - \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2(1-\alpha-\varepsilon)^2}\right)$$

where $\hat{I}_{n,\varepsilon} = \left[Y_{(i_\varepsilon)}, Y_{(j_\varepsilon)}\right]$

Consequence of Bennett's concentration inequality

Here a control on $\mathbb{E}_{\mathbf{P}_0}[L^2]$ is assumed. We denote $h(u) = (1+u)\log(1+u) - u$

Proposition

If $a \leq L \leq b$ and $\mathbb{E}_{P_0}[L^2] \leq \nu$ we have

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim P) \in \hat{I}_{n,\varepsilon}\right) \geq 1 - \exp\left(-\frac{n\nu_+}{a_+^2} h\left(\frac{a_+\varepsilon}{\nu_+}\right)\right) - \exp\left(-\frac{n\nu_-}{b_-^2} h\left(\frac{b_-\varepsilon}{\nu_-}\right)\right)$$

where a_+ and b_- are given by $a_+ = -a(-\alpha+\varepsilon)$ and $b_- = b(1-\alpha-\varepsilon)$, and ν_+ and ν_- are defined as

$$\nu_\pm = \min\{\nu, b\alpha\}(1-\alpha\pm\varepsilon) + \nu(\alpha^2+\varepsilon^2 \mp \alpha\varepsilon)$$

and $\hat{I}_{n,\varepsilon} = \left[Y_{(i_\varepsilon)}, Y_{(j_\varepsilon)}\right]$

Sub-Gaussian and sub-Gamma concentration inequalities

Here L is assumed light-tailed: **sub-Gaussian** or **sub-Gamma**. Denote $h_1(u) = u^2/2$ and $h_2(u) = 1 + u - \sqrt{1+2u}$

Proposition

If L is sub-Gaussian $(k = 1)$ with constant ν or sub-Gamma $(k = 2)$ with constants $\nu, c > 0$ then

$$\mathbb{P}\left(q_\alpha(Y|\boldsymbol{X} \sim P) \in \hat{I}_{n,\varepsilon}\right) \geq 1 - \exp\left(-\frac{n\nu_+}{c_+^2} h_k\left(\frac{c_+\varepsilon}{\nu_+}\right)\right) - \exp\left(-\frac{n\nu_-}{c_-^2} h_k\left(\frac{c_-\varepsilon}{\nu_-}\right)\right)$$

where ν_+, ν_- are given by

$$\nu_+ = (\alpha - \varepsilon)^2 \nu \quad \text{and} \quad \nu_- = (1 - \alpha - \varepsilon)^2 \nu,$$

c_+, c_- are given by $c_+ = (\alpha - \varepsilon)c$, $c_- = (1 - \alpha - \varepsilon)c$ and $\hat{I}_{n,\varepsilon} = \left[Y_{(i_\varepsilon)}, Y_{(j_\varepsilon)}\right]$

Tips for practionners

Hoeffding	Bennett	sub-Gaussian	sub-Gamma
$a \leq L \leq b$	$a \leq L \leq b$ and $\mathbb{E}_{P_0}[L^2] \leq \nu$	$\mathbb{E}_{P_0}[e^{\lambda L}] \leq e^{\lambda^2 \nu/2}$	$\mathbb{E}_{P_0}[e^{\lambda L}] \leq e^{\lambda^2 \nu/2(1-c\lambda)}$

Table: Summary of assumptions for each inequality

1. Hoeffding $\implies$ sub-Gaussian with $\nu = (b-a)^2/4$
2. Bennett is better than Hoeffding if $\mathbb{E}_{P_0}[L^2] = \nu \ll b - a$
3. sub-Gaussian with ν always better than sub-Gamma with ν (and any c)
4. L is heavy-tailed (i.e. $\mathbb{E}_{P_0}[L^m] = \infty$) $\implies$ no exponential bounds can be obtained

Back to the flood model

The **height** H of a river is computed using the previously defined model. We study **robustness** of

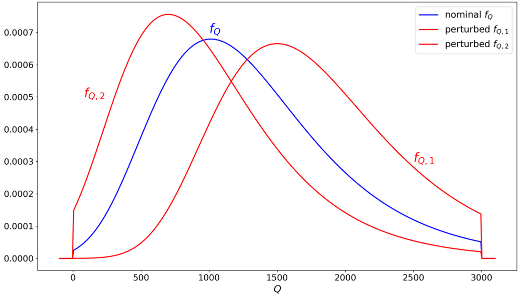
$$f_Q \longmapsto q_\alpha(H|Q \sim f_Q)$$

w.r.t. perturbations of f_Q to $f_{Q,1}$ and $f_{Q,2}$

Using an iid sample from $\mathbf{P}_0 = f_K \otimes f_Q \otimes f_{Z_m} \otimes f_{Z_v}$, we estimate and build CIs for

$$q_\alpha(H|Q \sim f_{Q,i})$$

- $f_Q, f_{Q,1}, f_{Q,2}$ are truncated Gumbel distributions
- $K \sim f_K$ truncated normal
- $Z_m, Z_v \sim f_{Z_m}, f_{Z_v}$ triangulars



Back to the flood model

Determining the constants:

- $a \leq \mathbf{L} = f_{Q,1}/f_Q \leq b$ where $a \approx 0.004$ and $b \approx 2.69$
- $\nu = \mathbb{E}_{\mathbf{P}_0}[\mathbf{L}^2] \approx 1.49$

	estimation	Hoeffding	Bennett
$\alpha = 0.5$	2.86m	$[2.83, 2.92]$	$[2.80, 2.95]$
$\alpha = 0.75$	3.42m	$[3.33, 3.50]$	$[3.30, 3.54]$
$\alpha = 0.95$	4.29m	$[3.97, 5.31]$	$[3.95, 5.45]$

Table: Estimation and 95% **non-asymptotic** confidence intervals for $q_\alpha(H|Q \sim f_{Q,1})$ with a sample of size **n=5000**

Back to the flood model

Determining the constants:

- $a \leq \boldsymbol{L} = f_{Q,2}/f_Q \leq b$ where $a \approx 0.40$ and $b \approx 5.93$
- $\nu = \mathbb{E}_{\mathbf{P_0}}[\boldsymbol{L}^2] \approx 1.38$

	estimation	Hoeffding	Bennett
$\alpha = 0.5$	1.99m	$[1.88, 2.10]$	$[1.90, 2.08]$
$\alpha = 0.75$	2.56m	$[2.39, 2.77]$	$[2.45, 2.68]$
$\alpha = 0.95$	3.54m	$[2.98, \infty)$	$[3.18, 4.52]$

Table: Estimation and 95% **non-asymptotic** confidence intervals for $q_\alpha(H|Q \sim f_{Q,2})$ with a sample of size **n=5000**

Back to the flood model

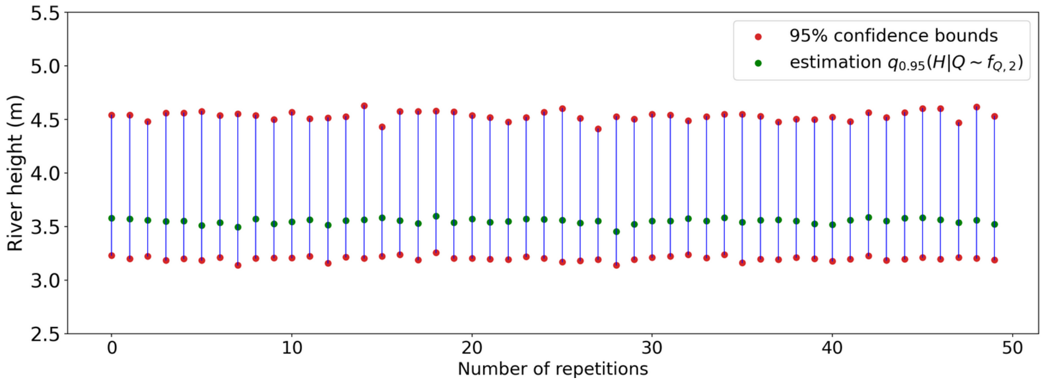


Figure: 50 estimations and 95% **non-asymptotic** CIs for $q_{0.95}(H|Q \sim f_{Q,2})$ with a sample of size **n=5000** using Bennett's inequality

Industrial safety application

We look at the **peak cladding temperature** T after a LOCA computed using CATHARE $T = G(X_1, \ldots, X_{56})$. We study **robustness** of

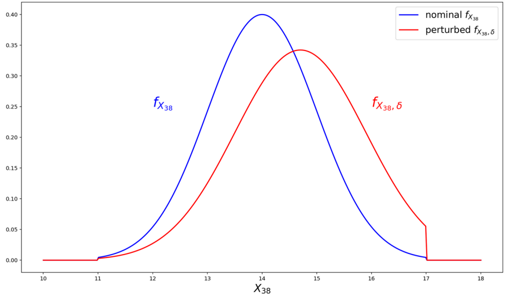
$$f_{X_{38}} \longmapsto q_\alpha(T | X_{38} \sim f_{X_{38}})$$

w.r.t. perturbations of $f_{X_{38}}$ to $f_{X_{38},\delta}$

Using an iid sample from $\mathbf{P}_0 = f_{X_1} \otimes \ldots \otimes f_{X_{56}}$, we estimate and build CIs for

$$q_\alpha(T | X_{38} \sim f_{X_{38},\delta})$$

- $f_{X_{38}}$ and $f_{X_{38},\delta}$ are truncated normals
- the other f_{X_i} are either truncated normal, truncated log-normal, uniform or log-uniform



Industrial safety application

Determining the constants:

- $a \leq L = f_{X_{38},\delta}/f_{X_{38}} \leq b$ where $a \approx 0.49$ and $b \approx 12.2$
- $\nu = \mathbb{E}_{\mathbf{P}_0}[L^2] \approx 1.76$

	estimation	Hoeffding	Bennett
$\alpha = 0.5$	299.5°C	$[292.4, 307.6]$	$[296.1, 303.1]$
$\alpha = 0.75$	328.5°C	$[310.3, 355.6]$	$[320.0, 337.8]$
$\alpha = 0.95$	382.4°C	$[336.0, \infty)$	$[362.1, 661.5]$

Table: Estimation and 95% **non-asymptotic** confidence intervals for $q_\alpha(T|X_{38} \sim f_{X_{38},\delta})$ with a sample of size **n=5000**

Industrial safety application

Determining the constants:

- $a \leq L = f_{X_{38},\delta}/f_{X_{38}} \leq b$ where $a \approx 0.49$ and $b \approx 12.2$
- $\nu = \mathbb{E}_{\mathbf{P}_0}[L^2] \approx 1.76$

	estimation ($n = 10^4$)	Hoeffding ($n = 5000$)	Hoeffding ($n = 10^4$)	Bennett ($n = 5000$)	Bennett ($n = 10^4$)
$\alpha = 0.5$	300.1°C	$[292.4, 307.6]$	$[294.7, 305.9]$	$[296.0, 303.1]$	$[297.4, 302.6]$
$\alpha = 0.75$	329.9°C	$[310.3, 355.6]$	$[314.5, 348.8]$	$[319.9, 337.8]$	$[323.3, 336.0]$
$\alpha = 0.95$	382.3°C	$[336.0, \infty)$	$[345.7, \infty)$	$[362.0, 661.5]$	$[368.0, 414.9]$

Table: Estimation and 95% **non-asymptotic** confidence intervals for $q_\alpha(T|X_{38} \sim f_{X_{38},\delta})$ with a sample of size **n=5000** and **n=10000**

Final remarks

Conclusion	Perspectives
■ Presented the **mathematical framework** for robustness analysis methods ■ Illustrated the **Fisher-Rao** perturbation method on parametric families ■ Built estimators and **non-asymptotic** CIs for quantiles $q_\alpha(Y \mid \boldsymbol{X} \sim \mathrm{P})$ generalizing Wilks method (preprint in preparation)	■ **Optimization algorithms** for computing robustness indices ■ Advising Pierre Schatz (internship) on the Fisher-Rao perturbation method for **dependent inputs** ■ **Maximum level** $\delta_{\max}$ for Fisher-Rao perturbation beyond which RA is not necessary ■ ...

References I

- Fang Chu and Marvin K Nakayama, *Confidence intervals for quantiles and value-at-risk when applying importance sampling*, Proceedings of the 2010 Winter Simulation Conference, IEEE, 2010, pp. 2751–2761.
- Tim Davie, *Fundamentals of hydrology*, Routledge, 2019.
- Herbert A David and Haikady N Nagaraja, *Order statistics*, John Wiley & Sons, 2004.
- Alvaro Rollon de Pinedo, *Statistical analysis of the results of numerical simulations of accidental situations in pressurized water reactors*, Ph.D. thesis, Université Grenoble Alpes [2020-....], 2021.
- Sébastien Da Veiga, Fabrice Gamboa, Bertrand Iooss, and Clémentine Prieur, *Basics and trends in sensitivity analysis: Theory and practice in r*, SIAM, 2021.

References II

- P. W. Glynn, *Importance sampling for monte carlo estimation of quantiles*, Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation, Citeseer, 1996, pp. 180–185.
- C. Gauchy, J. Stenger, R. Sueur, and B. Iooss, *An information geometry approach to robustness analysis for the uncertainty quantification of computer codes*, Technometrics **64** (2022), no. 1, 80–91.
- Marouane Il Idrissi, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, and Jean-Michel Loubes, *Quantile-constrained wasserstein projections for robust interpretability of numerical and machine learning models*, Electronic Journal of Statistics **18** (2024), no. 2, 2721–2770.

References III

B. B. Ketema, N. Bousquet, F. Costantino, F. Gamboa, B. Iooss, and R. Sueur, *Fisher-Rao distance between truncated distributions and robustness analysis in uncertainty quantification*, arXiv preprint arXiv:2407.21542 (2024).

Paul Lemaître, Ekatarina Sergienko, Aurélie Arnaud, Nicolas Bousquet, Fabrice Gamboa, and Bertrand Iooss, *Density modification-based reliability sensitivity analysis*, Journal of Statistical Computation and Simulation **85** (2015), no. 6, 1200–1223.

Jerome Stenger, *Optimal uncertainty quantification of a risk measurement from a computer code*, Ph.D. thesis, Université Paul Sabatier-Toulouse III, 2020.

Thank you for the attention !

Fundamental lemma

Lemma

Under continuity of F, the desired probability can be rewritten as

$$\mathbb{P}(Y_{(i_\xi)} \le q_\alpha \le Y_{(j_\xi)}) = 1 - \mathbb{P}\left(\sum_{i=1}^{n} Z_{i,\xi}^{+} < 0\right) - \mathbb{P}\left(\sum_{i=1}^{n} Z_{i,\xi}^{-} > 0\right), \tag{1}$$

where $(Z_{i,\xi}^{+})_{1\le i\le n}$ and $(Z_{i,\xi}^{-})_{1\le i\le n}$ are iid random variables each defined as

$$Z_{i,\xi}^{+} = L(Y_i)(\mathbf{1}_{Y_i \le q_\alpha} - \alpha + \xi) \quad \textit{and} \quad Z_{i,\xi}^{-} = L(Y_i)(\mathbf{1}_{Y_i \le q_\alpha} - \alpha - \xi).$$

In addition, the variables $Z_{i,\xi}^{\pm}$ have mean $\pm\xi$.

Contribution: application to truncated distributions

Engineers work with **bounded domains** $\Longrightarrow$ need to compute the Fisher-Rao distance on **truncated parametric families** i.e.

$$f_{\theta,ab}(x) = \frac{f_\theta(x)}{\int_a^b f_\theta(y)dy}\mathbf{1}_{x\in[a,b]}$$

How to do that ?

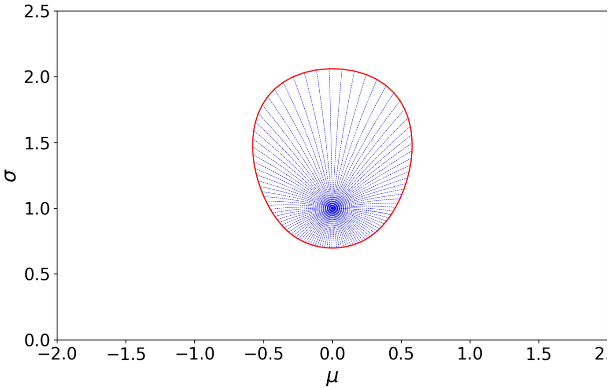
1. Compute the Fisher information (either explicitly or numerically),
2. Numerically solve an **ordinary differential equation** with coefficients involving the Fisher information

We considered a few well-known families: truncated normal, truncated log-normal, truncated Gumbel, Beta, triangular,...

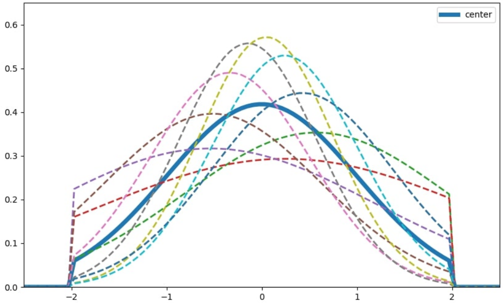
Preprint [KBC$^+$24] submitted and under review (arXiv: 2407.21542)

Contribution: application to truncated distributions

Perturbations in the **truncated normal** family



Sphere centered at $(0, 1)$ for $\delta = 0.4$

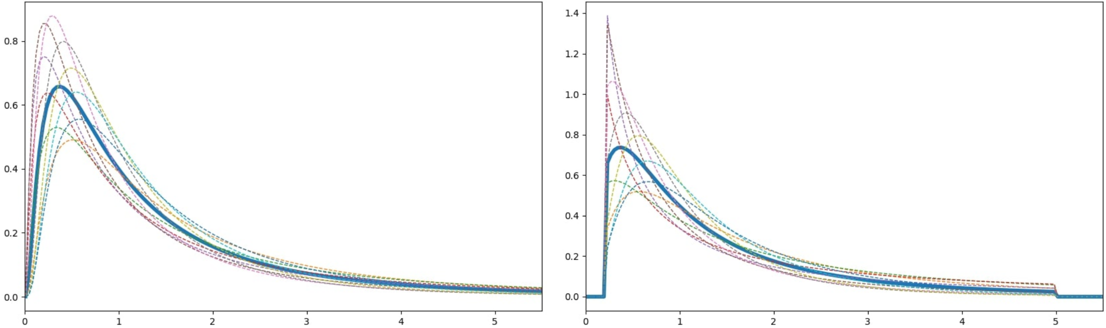


Perturbations of a truncated $\mathcal{N}(0, 1)$ where $\delta = 0.4$

Figure: **Truncated normal** family on $[-2, 2]$

Contribution: application to truncated distributions

Similarly for the truncated **log-normal** family (the Fisher information is the same as the truncated normal family)

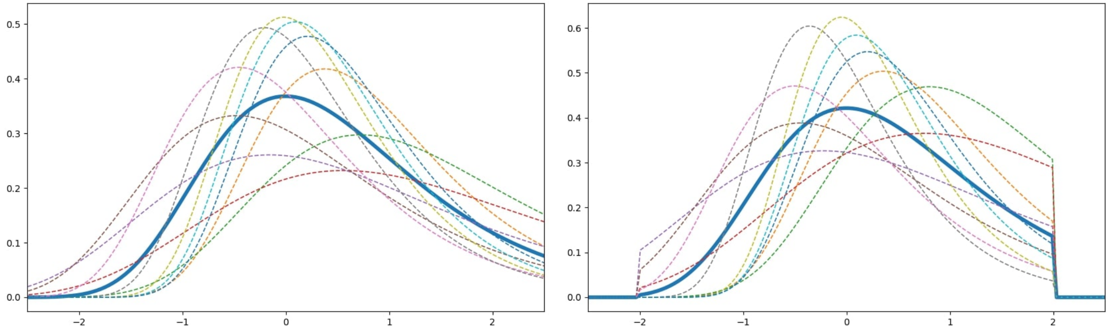


Perturbations ($\delta = 0.3$) of $\log \mathcal{N}(0,1)$ — Perturbations ($\delta = 0.3$) of truncated $\log \mathcal{N}(0,1)$

Figure: Perturbations of a **log-normal** density in the **usual** and **truncated** case

Contribution: application to truncated distributions

For the truncated **Gumbel** family



Perturbations ($\delta = 0.5$) of $\mathrm{Gumb}(0, 1)$ Perturbations ($\delta = 0.5$) of truncated $\mathrm{Gumb}(0, 1)$

Figure: Perturbations of a **Gumbel density** in the **usual** and **truncated** case