# Supplementary Material
# Augmented quantization: mixture models for risk-oriented sensitivity analysis

Charlie Sire [*][1], Didier Rullière[2], Rodolphe Le Riche[2], Jérémy Rohmer[3], Yann Richet[4], and Lucie Pheulpin[4]

[1]Mines Paris, PSL University, Centre for geosciences and geoengineering, Fontainebleau, France
[2]Mines Saint-Etienne, Univ. Clermont Auvergne, CNRS, UMR 6158 LIMOS
[3]BRGM
[4]IRSN

## 1   Augmented Quantization without perturbation

Here we perform AQ without the clusters perturbation, on a one-dimensional sample $S_u = (\mathbf{X}_i)_{i=1}^{300} \in \mathbb{R}^{300}$ to represent the uniform mixture of density $f = \frac{1}{3}\mathbb{1}_{[0,1]} + \frac{2}{3}\frac{\mathbb{1}_{[0.3,0.6]}}{0.3}$. Here, we start from two representatives $R_1 = R_U(0,0.5)$ and $R_2 = R_U(0.5,1)$, and Figure 1 shows the clusters obtained after each of the first 9 iterations. The algorithm is rapidly stuck in a configuration far from the optimal quantization, with representatives around $R_U(0.26,0.93)$ and $R_U(0.19,0.63)$.

The obtained quantization error here is $4.0 \times 10^{-2}$, almost 10 times higher than the optimal quantization error obtained with Augmented Quantization including clusters perturbation, equal to $4.4 \times 10^{-3}$.

## 2   Proof of proposition 3

This document provides the proof of the Proposition 3 of the associated manuscript, that shows the asymptotic clustering consistency:

**Proposition 3** (Asymptotic clustering consistency.). *If $R_j \in \mathcal{R}_s$ for $j \in \mathcal{J}$ and $\mathbf{X}_i \sim R_{J_i}$, $i = 1,\ldots,n$ with $(J_i)_{i=1}^n$ i.i.d. sample with same distribution as $J$, then*

$$\lim_{n,N \to +\infty} \mathbb{E}\left(\mathcal{E}_p(\boldsymbol{C}^\star(\boldsymbol{R},J,n,N))\right) = 0.$$

We recall and complete our notations used to discuss clustering:

- The representative mixtures are $(R_1,\ldots,R_\ell) \in \mathcal{R}_s^\ell$

---

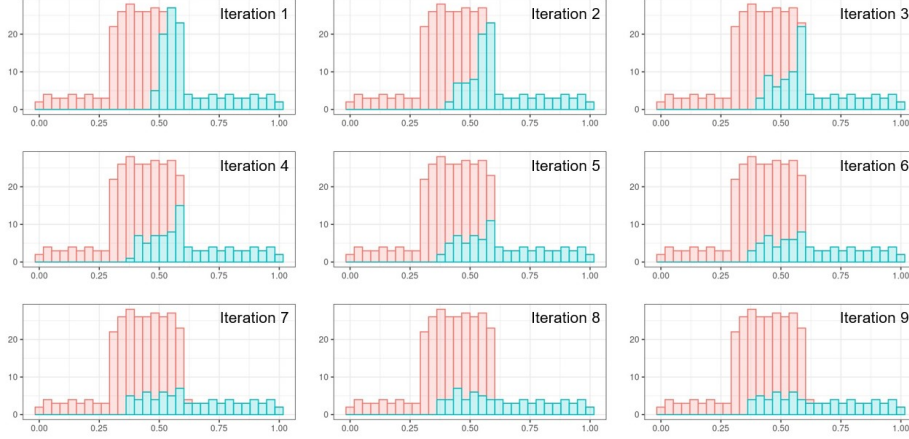[*]Corresponding author: `charlie.sire@minesparis.psl.eu`

Figure 1: Distribution of clusters 1 and 2 in the 1D uniform case when performing the method without the clusters perturbation.

- $J$ is a random variable with $\forall j \in \mathcal{J}$, $P(J = j) = p_j$

- $(\mathbf{X}_k)_{k=1}^{n}$ are i.i.d. data samples and $\mathbf{X}_k$ has probability measure $R_J$

- $(J_i)_{i=1}^{N}$ are i.i.d. samples with same distribution as $J$

- $(\mathbf{Y}_i)_{i=1}^{N}$ are i.i.d. mixture samples and $\mathbf{Y}_i$ has probability measure $R_{J_i}$

- The index of the representative whose sample is closest to $\mathbf{x}$, $I_N(\mathbf{x}) = \underset{i=1,\dots,N}{\arg\min} \|\mathbf{x} - \mathbf{Y}_i\|$

- The mixture sample the closest to $\mathbf{x}$, $a_N(\mathbf{x}) = \mathbf{Y}_{I_N(\mathbf{x})}$
  $$= \arg\min_{\mathbf{y} \in \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}} \|\mathbf{x} - \mathbf{y}\|$$

- The $k$-th data sample that is assigned to cluster $j$, $\tilde{\mathbf{X}}_{k,N}^{(j)} = \mathbf{X}_k \mid J_{I_N(\mathbf{X}_k)} = j$. This of course does not mean that the $k$-th sample will belong to cluster $j$, which may or may not happen. It is a random vector whose probability to belong to cluster $j$ will be discussed.

The space of measures of the representatives is $\mathcal{R}_s := \{\beta_{\mathrm{c}} R_{\mathrm{c}} + \beta_{\mathrm{disc}} R_{\mathrm{disc}}, R_{\mathrm{c}} \in \mathcal{R}_{\mathrm{c}}, R_{\mathrm{disc}} \in \mathcal{R}_{\mathrm{disc}}, \beta_{\mathrm{c}} + \beta_{\mathrm{disc}} = 1\}$ where

- $\mathcal{R}_{\mathrm{c}}$ is the set of the probability measures $R_{\mathrm{c}}$ for which the associated density (i.e. their Radon–Nikodým derivative with respect to the Lebesgue measure), exists, has finite support and is continuous almost everywhere in $\mathbb{R}^m$,

- $\mathcal{R}_{\mathrm{disc}}$ is the set of the probability measures of the form $R_{\mathrm{disc}} = \sum_{k=1}^{t} c_k \delta_{\gamma_k}$, with $\delta_{\gamma_k}$ the Dirac measure at $\gamma_k$ and $c_k \in ]0,1]$ such that $\sum_{k=1}^{t} c_k = 1$.

As $\mathcal{R}_s$ is stable under linear combinations, it follows that $R_J \in \mathcal{R}_s$, and it can be written $R_J = \beta_{\mathrm{c}} R_{\mathrm{c}} + \beta_{\mathrm{disc}} \sum_{k=1}^{t} c_k \delta_{\gamma_k}$. We denote by $f_c$ the probability density function (PDF) associated with $R_c$, and we assume that $\beta_{\mathrm{disc}} > 0$ and $\beta_{\mathrm{c}} > 0$, since the reasoning is simpler when either of them is zero.

2

## 2.1 Convergence in law of the mixture samples designated by the clustering

### 2.1.1 Convergence in law of $a_N(\mathbf{X}_k)$

The law of the closest sample from the mixture to the $k$-th data sample is similar, when $N$ grows, to that of any representative, for example the first one.

**Proposition 4.** $\forall k = 1, \ldots, n$, $a_N(\mathbf{X}_k)$ converges in law to $\mathbf{Y}_1$:

$$a_N(\mathbf{X}_k) \xrightarrow{\mathcal{L}} \mathbf{Y}_1$$

*Proof.* Let $\boldsymbol{\alpha} \in \mathbb{R}^m$. We denote $A(\boldsymbol{\alpha}) = \{\mathbf{y} \in \mathbb{R}^m, \forall i = 1, \ldots, m, \ y_i \leq \alpha_i\}$. The cumulated density function (CDF) of $a_N(\mathbf{X}_k)$ is

$$F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha}) =$$

$$\beta_c \int_{\mathbb{R}^m} P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) d\mathbf{x}$$

$$+ \beta_{\text{disc}} \sum_{r=1}^{t} c_r P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \gamma_r) \ .$$

The proof works by showing that the CDF of $a_N(\mathbf{X}_k)$ tends to that of $\mathbf{Y}_1$ when $N$ grows.

**Limit of** $\beta_{\text{disc}} \sum_{r=1}^{t} c_r P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \gamma_r)$
If $\gamma_r \in A(\boldsymbol{\alpha})$, then $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \gamma_r) \geq P(\exists i \in \{1, \ldots, N\}, \mathbf{Y}_i = \gamma_r)$. As a result, $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_r) \xrightarrow[N \to +\infty]{} 1$.
Else, if $\gamma_r \notin A(\boldsymbol{\alpha})$, then $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_r) \leq P(\forall i \in \{1, \ldots, N\}, \mathbf{Y}_i \neq \gamma_r)$. Consequently, $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_k) \xrightarrow[N \to +\infty]{} 0$.
Finally, it follows that $\beta_{\text{disc}} \sum_{r=1}^{m} c_r P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_r) \xrightarrow[N \to +\infty]{} \beta_{\text{disc}} \sum_{r \ s.t \ \gamma_r \in A(\boldsymbol{\alpha})} c_r$

**Limit of** $\beta_c \int_{\mathbb{R}^m} P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(x) dx$

Let $\mathbf{x} \in \mathbb{R}^m$ such that $f_c$ is continuous at $\mathbf{x}$ and $\mathbf{x} \notin \{\mathbf{x} \in \mathbb{R}^m, \exists i \in \{1, \ldots, d\}, \ x_i = \alpha_i\}$.

- If $f_c(\mathbf{x}) = 0$ then $\forall N, P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) = 0$

- Case when $\mathbf{x} \in A(\boldsymbol{\alpha})$ and $f_c(\mathbf{x}) > 0$. Then, $\exists (\varepsilon_1, \ldots, \varepsilon_m) \in (\mathbb{R}_+^\star)^m, x_k = \alpha_k - \varepsilon_k$.
  By continuity, $\exists \eta \in ]0, \arg\min_{1 \leq k \leq m} \varepsilon^k[, \forall \mathbf{x}' \in \mathcal{B}(\mathbf{x}, \eta), f_c(\mathbf{x}') > 0$, with $\mathcal{B}(\mathbf{x}, \eta)$ the ball of radius $\eta$ and center $\mathbf{x}$. It leads to

$$P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) \geq P(\exists i \in \{1, \ldots, N\}, \mathbf{Y}_i \in \mathcal{B}(\mathbf{x}, \eta)) \xrightarrow[N \to +\infty]{} 1$$

- Case when $\mathbf{x} \notin A(\boldsymbol{\alpha})$ and $f_c(\mathbf{x}) > 0$. Then, $\exists \varepsilon > 0, x_k = \alpha_k + \varepsilon$. By continuity, $\exists \eta \in ]0, \varepsilon[$, $\forall \mathbf{x}' \in \mathcal{B}(\mathbf{x}, \eta), f_c(\mathbf{x}') > 0$, with $\mathcal{B}(\mathbf{x}, \eta)$ the ball of radius $\eta$ and center $\mathbf{x}$. Then,

$$P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) \leq P(\forall i \in \{1, \ldots, N\}, \mathbf{Y}_i \notin \mathcal{B}(\mathbf{x}, \eta)) \xrightarrow[N \to +\infty]{} 0.$$

We then have, $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) \xrightarrow[N \to +\infty]{a.e.} \mathbb{1}_{A(\boldsymbol{\alpha})}(\mathbf{x}) f_c(\mathbf{x})$.

The Dominated Convergence Theorem gives,

$$\beta_c \int_{\mathbb{R}^m} P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) d\mathbf{x}$$

$$\xrightarrow[N \to +\infty]{} \beta_c \int_{A(\boldsymbol{\alpha})} f_c(\mathbf{x}) d\mathbf{x}$$

**Limit of $F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha})$**

Finally, we have,

$$F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha}) \xrightarrow[N \to +\infty]{} F_Y(\boldsymbol{\alpha}) \ .$$

$\square$

### 2.1.2 Convergence in law of $a_N(\mathbf{X}_k) \mid J_{I_N(\mathbf{X}_k)} = j$

The random vector $a_N(\mathbf{X}_k) \mid J_{I_N(\mathbf{X}_k)} = j$, i.e., the closest mixture sample to the $k$-th data sample, when constrained to belong to the $j$-th cluster and when $N$ grows, follows the distribution of the $j$-th representative $R_j$.

**Proposition 5.**

$$\forall j \in \mathcal{J}, \ a_N(\mathbf{X}_k) \mid J_{I_N(\mathbf{X}_k)} = j \xrightarrow{\mathcal{L}} \mathbf{Y}_1 \mid J_1 = j$$

*Proof.* As in the previous Proposition, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ and $A(\boldsymbol{\alpha}) = \{\mathbf{x} \in \mathbb{R}^m, \forall i = 1, \ldots, m, \ x_i \leq \alpha_i\}$.

To shorten the notations, we write the $k$-th sample point associated to the $j$-th cluster $\tilde{\mathbf{X}}_{k,N}^{(j)} := \mathbf{X}_k \mid J_{I_N(\mathbf{X}_k)} = j$ .

$$F_{a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})}(\boldsymbol{\alpha})$$
$$= P(a_N(\tilde{\mathbf{X}}_{k,N}^{(j)}) \in A(\boldsymbol{\alpha}))$$
$$= P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid J_{I_N(\mathbf{X}_k)} = j)$$
$$= \sum_{i=1}^N P(I_N(\mathbf{X}_k) = i, \mathbf{Y}_i \in A(\boldsymbol{\alpha}) \mid J_{I_N(\mathbf{X}_k)} = j)$$
$$= N P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_{I_N(\mathbf{X}_k)} = j)$$

Bayes' Theorem gives,

$$P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_{I_N(\mathbf{X}_k)} = j)$$

$$= \frac{P(J_{I_N(\mathbf{X}_k)} = j \mid I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(J_{I_N(\mathbf{X}_k)} = j)}$$
$$\times P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))$$

$$= \frac{P(J_1 = j \mid \mathbf{Y}_1 \in A(\boldsymbol{\alpha})) P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(J_{I_N(\mathbf{X}_k)} = j)}$$

$$= \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) P(J_1 = j)}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}$$
$$\times \frac{P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(J_{I_N(\mathbf{X}_k)} = j)}$$

$$= \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) p_j}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))} \frac{P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{p_j}$$

$$= \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}$$

Then

$$F_{a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})}(\boldsymbol{\alpha})$$

$$= N \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) P(I_N(\mathbf{X}_k) = 1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}$$

$$= \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j)}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))} \sum_{i=1}^{N} P(I_N(\mathbf{X}_k) = i, \mathbf{Y}_i \in A(\boldsymbol{\alpha}))$$

$$= P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) \frac{P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}$$

Proposition 4 gives $\frac{P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))} \xrightarrow[N \to +\infty]{} 1.$

As a result,

$$F_{a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})}(\boldsymbol{\alpha}) \xrightarrow[N \to +\infty]{} P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1 = j) \ .$$

$\square$

## 2.2 Wasserstein convergence to the representatives

### 2.2.1 Wasserstein convergence of $a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})$

The mixture sample that is both the closest to $\mathbf{X}_k$ and coming from $R_j$ has a distribution that converges, in terms of Wasserstein metric, to $R_j$. This is a corollary of Proposition 5.

**Corollary 1.**

$$\mathcal{W}_p(\mu_N^j, R_j) \xrightarrow[N \to +\infty]{} 0,$$

*where $\mu_N^j$ is the measure associated to $a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})$.*

*Proof.* As stated in Villani, 2016, convergence in $p$-Wasserstein distance is equivalent to convergence in law plus convergence of the $p^{th}$ moment. As the distributions of $\mathcal{R}$ have bounded supports, $X \xrightarrow{\mathcal{L}} Y \implies E(\| X \|^p) = E(\| Y \|^p)$. $\square$

### 2.2.2 Wasserstein convergence of $(a_N(\tilde{\mathbf{X}}_{k,N}^{(j)}))_{k=1}^n$

The above Wasserstein convergence to the $j$-th representative for a given data sample $\mathbf{X}_k$ also works on the average of the $\mathbf{X}_k$'s as their number $n$ grows.

**Proposition 6.**
$$\lim_{n,N \to +\infty} \mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, R_j)) = 0,$$

*where $\mu_{n,N}^j$ is the empirical measure associated to $(a_N(\tilde{\mathbf{X}}_{k,N}^{(j)}))_{k=1}^n$.*

*Proof.* By the triangle inequality applied to the $\mathcal{W}_p(,)$ metric,

$$\mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, R_j)) \leq \mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, \mu_N^j)) +$$
$$\mathbb{E}(\mathcal{W}_p(\mu_N^j, R_j)) .$$

As the distributions of $\mathcal{R}_s$ have a bounded support, then for all $q \in \mathbb{N}$, $\exists M_1, \forall N \in \mathbb{N}, \int_{\mathcal{X}} |\mathbf{x}|^q \mu_N^j(d\mathbf{x}) \leq M_1$.
Then, as shown in Fournier and Guillin, 2013,

$$\exists M_2, \forall N \in \mathbb{N}, \ \mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, \mu_N^j)) \leq M_2 G(n)$$

where $G(n) \underset{n \to +\infty}{\longrightarrow} 0$.
Let $\varepsilon > 0$,

$$\exists n_0, \forall n > n_0, \forall N, \ \mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, \mu_N^j)) < \frac{\varepsilon}{2} ,$$

and Corollary 1 gives

$$\exists N_0, \forall N > N_0, \ \mathbb{E}(\mathcal{W}_p(\mu_{N,j}, R_j)) < \frac{\varepsilon}{2} .$$

Finally,

$$\forall n, N > max(n_0, N_0), \ \mathbb{E}(\mathcal{W}_p(\mu_{n,N}^j, R_j) < \varepsilon .$$

$\square$

### 2.2.3 Wasserstein convergence of $C_j^\star(\boldsymbol{R}, J, n, N)$ to $R_j$

The $j$-th cluster produced by the *FindC* algorithm, $C_j^\star(\boldsymbol{R}, J, n, N)$, can be seen as the i.i.d. sample $(\tilde{\mathbf{X}}_{k,N}^{(j)})_{k=1}^{n_j}$, with $\sum_{j=1}^\ell n_j = n$. We have $n_j \to +\infty$ when $n \to +\infty$, provided $p_j > 0$.

We denote $\nu_{n_j,N}^j$ the empirical measure associated to $(\tilde{\mathbf{X}}_{k,N}^{(j)})_{k=1}^{n_j}$.
The triangle inequality implies that

$$\mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, R_j)) \leq \mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, \mu_{n_j,N}^j)) +$$
$$\mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, R_j)) .$$

Furthermore,

$$\mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, \mu_{n_j,N}^j))$$
$$\leq \mathbb{E}(\frac{1}{n_j} \sum_{k=1}^{n_j} \|\tilde{\mathbf{X}}_{k,N}^{(j)} - a_N(\tilde{\mathbf{X}}_{k,N}^{(j)})\|^p)$$
$$\leq \mathbb{E}(\|\tilde{\mathbf{X}}_{1,N}^{(j)} - a_N(\tilde{\mathbf{X}}_{1,N}^{(j)})\|^p)$$
$$\leq \mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p | J_{I_N(\mathbf{X}_1)} = j)$$

The law of total expectation gives

$$\mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p)$$

$$= \sum_{j=1}^{\ell} p_j \mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p \mid J_{I_N(\mathbf{X}_1)} = j) \ .$$

Since $a_N(\mathbf{X}_k)$ is the closest sample in $\{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$ to $\mathbf{X}_k$, we have

$$\mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p)$$

$$= \mathbb{E}(\frac{1}{N} \sum_{k=1}^{N} \|\mathbf{X}_k - a_N(\mathbf{X}_k)\|^p)$$

$$\leq \mathbb{E}(\mathcal{W}_p((\mathbf{X}_k)_{k=1}^{N}, (\mathbf{Y}_i)_{i=1}^{N})) \xrightarrow[N \to +\infty]{} 0$$

The last two relations give,

$$\forall j \in \mathcal{J},$$
$$\mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p \mid J_{I_N(\mathbf{X}_1)} = j) \xrightarrow[N \to +\infty]{} 0$$

Therefore,

$$\begin{cases} \lim\limits_{n_j, N \to +\infty} \mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, \mu_{n_j,N}^j)) = 0 \ , \\ \lim\limits_{n_j, N \to +\infty} \mathbb{E}(\mathcal{W}_p(\nu_{n_j,N}^j, R_j)) = 0 \text{ by Proposition 6.} \end{cases}$$

## 2.3 Convergence of the expected quantization error

We have

$$\mathcal{E}_p(\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N))$$

$$= \left( \sum_{j \in \mathcal{J}} p_j w_p(C_j^\star(\boldsymbol{R}, J, n, N))^p \right)^{1/p}$$

$$\leq \sum_{j \in \mathcal{J}} p_j w_p(C_j^\star(\boldsymbol{R}, J, n, N))$$

where $w_p(C_j) := \mathcal{W}_p(C_j, R_j^\star(C_j)) = \min\limits_{r \in \mathcal{R}} \mathcal{W}_p(C_j, r)$, denotes the local error associated for a given cluster $C_j$. By definition of the local error, for $j \in \mathcal{J}$, $w_p(C_j^\star(\boldsymbol{R}, J, n, N)) \leq \mathcal{W}_p(C_j^\star(\boldsymbol{R}, J, n, N), R_j)$ . As just seen in Section 2.2.3,

$$\lim_{n_j, N \to +\infty} \mathbb{E}\left( \mathcal{W}_p(\nu_{n_j,N}^j, R_j) \right)$$

$$= \lim_{n, N \to +\infty} \mathbb{E}\left( \mathcal{W}_p(C_j^\star(\boldsymbol{R}, J, n, N), R_j) \right)$$

$$= 0 \ .$$

Then,

$$\lim_{n, N \to +\infty} \mathbb{E}\left( w_p(C_j^\star(\boldsymbol{R}, J, n, N)) \right) = 0 \ ,$$

and finally,

$$\lim_{n, N \to +\infty} \mathbb{E}\left( \mathcal{E}_p(\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N)) \right) = 0 \ .$$

# 3  *FindR* step for mixtures of Dirac and uniform measures

In this section we provide details about *FindR* step with the family $\mathcal{R}$ introduced in Section 4. This step consists of finding, in the sense of Proposition 2, the optimal parameters

$$(\alpha_1, \ldots, \alpha_m, a_1, \ldots, a_m, \sigma_1, \ldots, \sigma_m)$$

with $\alpha_k$ a boolean $\sigma_k = 0$ if and only if $\alpha_k = 0$.. The optimization is performed separately on each marginal, to avoid the computation of multivariate Wasserstein distances. Then, for each cluster $C_j$, we define here its $k$-th marginal: $C_j^k = \{x_k : (x_1, \ldots, x_m) \in C_j\}$, and we minimize

$$\mathcal{W}_2\left(C_j^k, R_k(\alpha_k, a_k, \sigma_k)\right),$$

where

- $R_k$ is a uniform between $a_k - \frac{\sigma_k}{2}$ and $a_k + \frac{\sigma_k}{2}$ if the boolean $\alpha_k = 1$, with $\sigma_k \in \{0.25, 0.5, 0.75, 1\}$.

- If $\alpha_k = 0$, $R_k$ is a Dirac at $a_k \in [0, 1]$, and $\sigma_k = 0$.

Then, the four possibilities are tested for the pair $(\alpha_k, \sigma_k)$, and we have the following analytical result:

- If $\alpha_k = 0$, then $\sigma_k = 0$ and $R_k$ is a Dirac measure at $a_k$. The optimal value for $a_k$ is

$$a_k^\star = \frac{1}{\text{card}(C_j)} \sum_{x \in C_j^k} x$$

  the barycenter of $C_j^k$.

- If $\alpha_k = 1$, then $\sigma_k \in \{0.25, 0.5, 0.75, 1\}$ and $R_k$ is a 1D-uniform measure with support center $a_k$ and support width $\sigma_k$. The optimal value for $a_k$ is

$$a_k^\star = \int_0^1 Q_j^k(q)(-6q + 4)dq,$$

  with $Q_j^k$ the empirical quantile function of $C_j^k$.

# 4  Screening on G-function

We consider a test case based on the G-function introduced in Sobol, 1998, defined as

$$G(\mathbf{X}) = \prod_{i=1}^d \frac{|4X_i - 2| + a_i}{1 + a_i},$$

where the input variables $X_i \sim \mathcal{U}(0, 1)$ are independent, $a_i \geq 0$ are coefficients controlling variable influence, and $d$ is the dimension. In this study, we set $d = 10$.

The coefficients $a_i$ determine the impact of each variable on the output: smaller values of $a_i$ correspond to stronger influence, whereas larger $a_i$ indicate weaker influence. The objective is to perform screening and identify the most influential variables on the risk event $G(\mathbf{X}) > q_{95}$, where $q_{95}$ is the 0.95-quantile of $G(\mathbf{X})$.. We do this using AQ by considering a mixture of distributions with independent marginals, where each marginal is either a Dirac delta or a uniform distribution with support width 1.
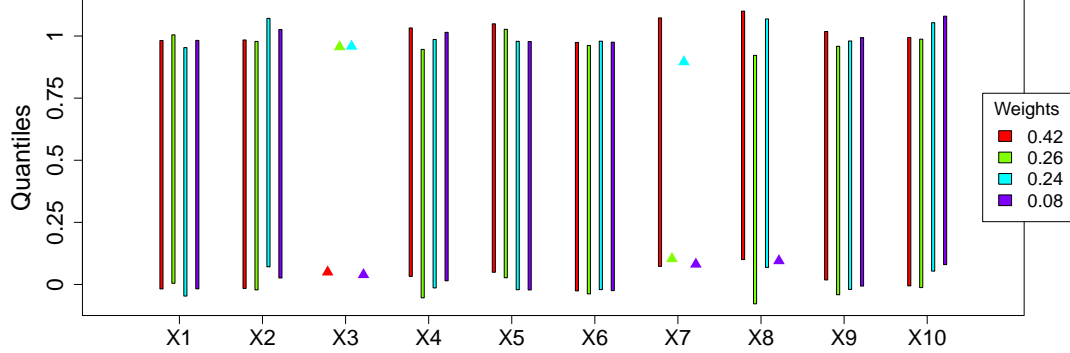
For this example, we choose:

Figure 2: Scenarios leading to $G(\mathbf{X}) > q_{95}$ in the flooding test case: mixture of four distributions with Dirac and large uniform independent marginals. Each distribution is associated with a color (red, green, blue and purple). The mixture weights are 0.42 for the red component, 0.26 for the green component, 0.24 for the blue component, and 0.08 for the purple one, as indicated on the right side of the plot. A vertical bar represents a uniform distribution, while a triangle marks the location of a Dirac.

- $a_i = 99$ for $i \in \{1, 2, 4, 5, 6, 9, 10\}$, making these variables essentially non-influential,

- $a_3 = 0$, giving $X_3$ the largest influence,

- $a_7 = 2$, giving $X_7$ a moderate influence,

- $a_8 = 4$, giving $X_8$ a small influence.

From independent samples $(\mathbf{X}_i)_{i=1}^{6000}$, we obtain $n = 281$ realizations of

$$\big(F_1(X_1), \ldots, F_{10}(X_{10})\big) \mid G(\mathbf{X}) > q_{95}.$$

Figure 2 provides the obtained mixture, that confirms the expectations given the chosen parametrization:

- For all variables $i \notin \{3, 7, 8\}$, the obtained mixture exhibits only large uniform representatives, indicating that these variables are non-influential, as expected.

- For $X_3$, only Dirac representatives are identified, showing its strong influence on $G(\mathbf{X}) > q_{95}$.

- Variable $X_7$ can take any value in 42% of the scenarios (red representative), while only specific values of this variable lead to $Y > q_{95}$ in the remaining cases (Dirac for the three other representatives). This confirms that its influence is smaller than that of $X_3$, but still significant.

- The influence of $X_8$ is highlighted in only 8% of the cases (purple scenario), confirming that although it still has an effect, this effect is smaller than for $X_3$ and $X_7$.

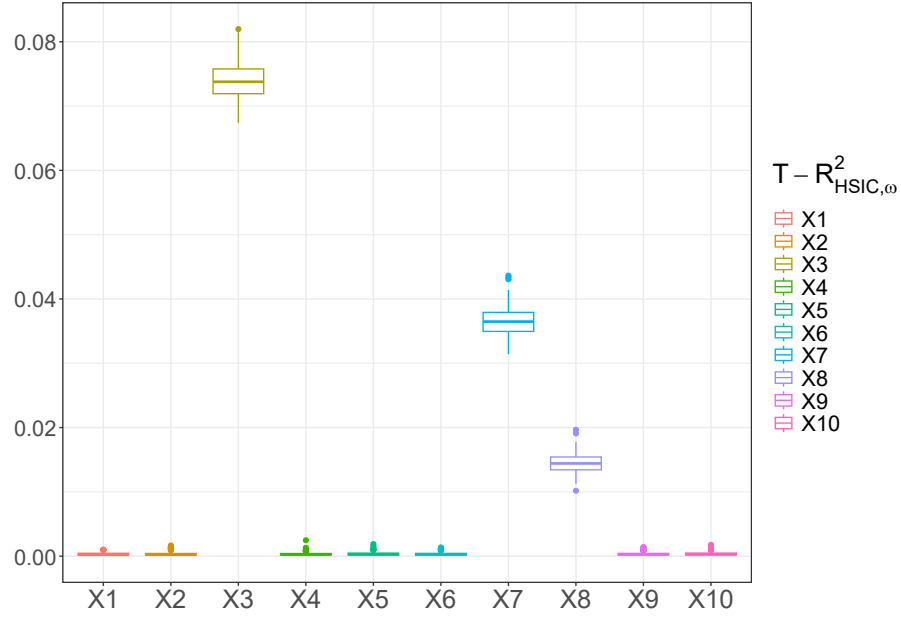This analysis is entirely confirmed by the target HSIC indices, provided in Figure 3.

9

Figure 3: Target HSIC T-$R^2_{\mathrm{HSIC},\omega}(X_i, G(\mathbf{X}))$ estimated from samples of size 6000 in the flooding test case, plotted on a logarithmic scale.

# References

Fournier, N., & Guillin, A. (2013). On the rate of convergence in wasserstein distance of the empirical measure.

Sobol, I. M. (1998). On quasi-monte carlo integrations. *Mathematics and computers in simulation*, *47*(2-5), 103–112.

Villani, C. (2016). *Optimal transport: Old and new*. Springer Berlin Heidelberg. https://books.google.fr/books?id=5p8SDAEACAAJ