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 4.0×10^{-2}, almost 10 times higher than the optimal quantization error obtained with Augmented Quantization including clusters perturbation, equal to 4.4×10^{-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 $\mathbf{X}_i \sim R_{J_i}$, $i = 1, \dots, n$ 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 $(R_1, \dots, R_\ell) \in \mathcal{R}_s^\ell$

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

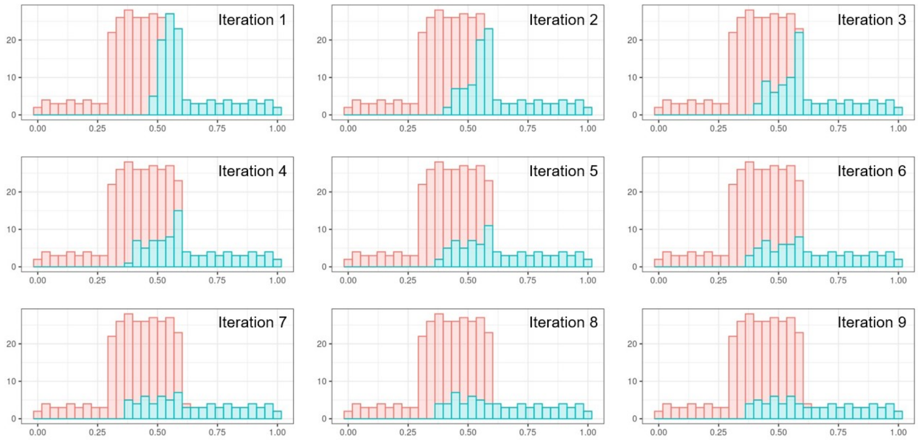

Figure 1: Distribution of clusters 1 and 2 in the 1D uniform case when performing the method without the clusters perturbation.

- 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_{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 δ_{γ_k} the Dirac measure at γ_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.

2.1 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

$$\begin{aligned}
&F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha}) = \\
&\beta_c \int_{\mathbb{R}^m} P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \mathbf{x}) f_c(\mathbf{x}) d\mathbf{x} \\
&+ \beta_{\text{disc}} \sum_{r=1}^{t} c_r P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X}_k = \gamma_r)\ .
\end{aligned}$$

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, $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_r) \underset{N \to +\infty}{\longrightarrow} 1$.

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, $P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_k) \underset{N \to +\infty}{\longrightarrow} 0$.

Finally, it follows that $\beta_{\text{disc}} \sum_{r=1}^{m} c_r P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}) \mid \mathbf{X} = \gamma_r) \underset{N \to +\infty}{\longrightarrow} \beta_{\text{disc}} \displaystyle\sum_{r\ s.t\ \gamma_r \in A(\boldsymbol{\alpha})} c_r$

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, $\exists \eta \in]0, \underset{1 \leq k \leq m}{\arg\min}\, \varepsilon^k[$, $\forall \mathbf{x}' \in \mathcal{B}(\mathbf{x}, \eta), f_c(\mathbf{x}') > 0$, with $\mathcal{B}(\mathbf{x}, \eta)$ the ball of radius η and center $\mathbf{x}$. It leads to

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

- Case when $\mathbf{x} \notin A(\boldsymbol{\alpha})$ and $f_c(\mathbf{x}) > 0$. Then, $\exists \varepsilon > 0, \mathbf{x}_k = \alpha_k + \varepsilon$. 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 η and center $\mathbf{x}$. Then,

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

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

The Dominated Convergence Theorem gives,

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

Limit of $F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha})$
Finally, we have,

$$F_{a_N(\mathbf{X}_k)}(\boldsymbol{\alpha}) \underset{N\to+\infty}{\longrightarrow} F_Y(\boldsymbol{\alpha}) \ .$$

□

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, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$ and $A(\boldsymbol{\alpha}) = \{\mathbf{x} \in \mathbb{R}^m, \forall i = 1, \dots, m,\ x_i \leq \alpha_i\}$.
To shorten the notations, we write the k-th sample point associated to the j-th cluster $\tilde{\mathbf{X}}_{k,N}^{(j)} \coloneqq \mathbf{X}_k \mid J_{I_N(\mathbf{X}_k)} = j$.

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

Bayes' Theorem gives,

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

Then

$$
\begin{aligned}
&F_{a_N(\tilde{\mathbf{X}}^{(j)}_{k,N})}(\boldsymbol{\alpha}) \\
&= N\frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1=j)P(I_N(\mathbf{X}_k)=1, \mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))} \\
&= \frac{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1=j)}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))} \sum_{i=1}^{N} P(I_N(\mathbf{X}_k)=i, \mathbf{Y}_i \in A(\boldsymbol{\alpha})) \\
&= P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}) \mid J_1=j)\frac{P(a_N(\mathbf{X}_k) \in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1 \in A(\boldsymbol{\alpha}))}
\end{aligned}
$$

Proposition 4 gives $\frac{P(a_N(\mathbf{X}_k)\in A(\boldsymbol{\alpha}))}{P(\mathbf{Y}_1\in A(\boldsymbol{\alpha}))} \underset{N\to+\infty}{\longrightarrow} 1$.
As a result,

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

□

2.2 Wasserstein convergence to the representatives

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

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.

$$
\mathcal{W}_p(\mu^j_N, R_j) \underset{N\to+\infty}{\longrightarrow} 0,
$$

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

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, $X \xrightarrow{\mathcal{L}} Y \Longrightarrow E(\| X \|^p) = E(\| Y \|^p)$. □

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

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.

$$\underset{n,N\to+\infty}{lim} \mathbb{E}(\mathcal{W}_p(\mu^j_{n,N}, R_j)) = 0,$$

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

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

$$\begin{aligned}\mathbb{E}(\mathcal{W}_p(\mu^j_{n,N}, R_j)) \leq & \mathbb{E}(\mathcal{W}_p(\mu^j_{n,N}, \mu^j_N)) + \\ & \mathbb{E}(\mathcal{W}_p(\mu^j_N, R_j)) \ .\end{aligned}$$

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

$$\exists M_2, \forall N \in \mathbb{N},\ \mathbb{E}(\mathcal{W}_p(\mu^j_{n,N}, \mu^j_N)) \leq M_2 G(n)$$

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

$$\exists n_0, \forall n > n_0, \forall N,\ \mathbb{E}(\mathcal{W}_p(\mu^j_{n,N}, \mu^j_N)) < \frac{\varepsilon}{2} \ ,$$

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,

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

□

2.2.3 Wasserstein convergence of $C^\star_j(\boldsymbol{R}, J, n, N)$ to R_j

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

We denote $\nu^j_{n_j,N}$ the empirical measure associated to $(\tilde{\mathbf{X}}^{(j)}_{k,N})^{n_j}_{k=1}$.

The triangle inequality implies that

$$\begin{aligned}\mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, R_j)) \leq & \mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, \mu^j_{n_j,N})) + \\ & \mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, R_j)) \ .\end{aligned}$$

Furthermore,

$$\begin{aligned}& \mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, \mu^j_{n_j,N})) \\ & \leq \mathbb{E}(\frac{1}{n_j} \sum_{k=1}^{n_j} \|\tilde{\mathbf{X}}^{(j)}_{k,N} - a_N(\tilde{\mathbf{X}}^{(j)}_{k,N})\|^p) \\ & \leq \mathbb{E}(\|\tilde{\mathbf{X}}^{(j)}_{1,N} - a_N(\tilde{\mathbf{X}}^{(j)}_{1,N})\|^p) \\ & \leq \mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p|\ J_{I_N(\mathbf{X}_1)} = j)\end{aligned}$$

The law of total expectation gives

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

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

$$\begin{aligned}&\mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p) \\ &= \mathbb{E}(\frac{1}{N}\sum_{k=1}^{N}\|\mathbf{X}_k - a_N(\mathbf{X}_k)\|^p) \\ &\leq \mathbb{E}(\mathcal{W}_p((\mathbf{X}_k)_{k=1}^N, (\mathbf{Y}_i)_{i=1}^N)) \underset{N\to+\infty}{\longrightarrow} 0\end{aligned}$$

The last two relations give,

$$\begin{aligned}&\forall j \in \mathcal{J}, \\ &\quad \mathbb{E}(\|\mathbf{X}_1 - a_N(\mathbf{X}_1)\|^p |\ J_{I_N(\mathbf{X}_1)} = j) \underset{N\to+\infty}{\longrightarrow} 0\end{aligned}$$

Therefore,

$$\begin{cases} \underset{n_j, N\to+\infty}{lim} \mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, \mu^j_{n_j,N})) = 0\ , \\ \underset{n_j, N\to+\infty}{lim} \mathbb{E}(\mathcal{W}_p(\nu^j_{n_j,N}, R_j)) = 0 \text{ by Proposition 6.} \end{cases}$$

2.3 Convergence of the expected quantization error

We have

$$\begin{aligned}&\mathcal{E}_p(\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N)) \\ &= \left(\sum_{j\in\mathcal{J}} p_j w_p(C^\star_j(\boldsymbol{R}, J, n, N))^p\right)^{1/p} \\ &\leq \sum_{j\in\mathcal{J}} p_j w_p(C^\star_j(\boldsymbol{R}, J, n, N))\end{aligned}$$

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

$$\begin{aligned}&\underset{n_j, N\to+\infty}{lim} \mathbb{E}\left(\mathcal{W}_p(\nu^j_{n_j,N}, R_j)\right) \\ &= \underset{n, N\to+\infty}{lim} \mathbb{E}\left(\mathcal{W}_p(C^\star_j(\boldsymbol{R}, J, n, N), R_j)\right) \\ &= 0\ .\end{aligned}$$

Then,

$$\underset{n, N\to+\infty}{lim} \mathbb{E}\left(w_p(C^\star_j(\boldsymbol{R}, J, n, N))\right) = 0\ ,$$

and finally,

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

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

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

with α_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: $C_j^k = \{x_k : (x_1, \dots, x_m) \in C_j\}$, 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 (α_k, σ_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

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

  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 σ_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.

References

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

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