# Augmented Quantization: a General Approach to Mixture Models

Charlie Sire, Rodolphe Le Riche, Didier Rullière, Jérémy Rohmer, Lucie Pheulpin, Yann Richet

# Augmented Quantization: a General Approach to Mixture Models

Charlie SIRE[1]

Supervisors: R. LE RICHE[3], D. RULLIERE[3], J. ROHMER[2], L. PHEULPIN[4], Y. RICHET[4]

[1]Inria Saclay - Ecole Polytechnique

[2]BRGM

[3]Mines Saint-Etienne and CNRS,LIMOS

[4]IRSN

# Content

# Mixture models

Consider $\mathcal{R}$ a family of probability measures, $\ell \in \mathbb{N}^\star$ and $\mathcal{J} = \{1, \dots, \ell\}$.

**Objective:** Approximate the distribution of $(x_i)_{i=1}^n$ by the mixture $R_J$ with

- $R_J = \sum_{j \in \mathcal{J}} p_j R_j$ ,
- $R_j \in \mathcal{R}, \quad j \in \mathcal{J}$ ,
- $J$ a discrete random variable with weights $p_j = P(J = j)$, $j \in \mathcal{J}$.

# EM methods

Classical approaches for mixture models use the notion of likelihood (Dellaert 2003, Sridharan 2014, Delyon et al. 1999, McLachlan et al. 2019, Nguyen et al. 2020) but they are not adapted to all distributions:

- Problem of definition for Dirac distributions
- Problem of support for uniform distributions

$\Longrightarrow$ Introduce an approach based on the quantization problem

# Quantization error with Wasserstein

Sample $(x_i)_{i=1}^n \in \mathcal{X}^n$

Principle: Find $\Gamma_\ell = (\gamma_1, \dots, \gamma_\ell) \in \mathcal{X}^\ell$ minimizing

$$\mathcal{E}_p(\Gamma_\ell) := \left( \frac{1}{n} \sum_{i=1}^{n} \|x_i - \arg\min_{\gamma \in \Gamma_\ell} \|x_i - \gamma\|\|^p \right)^{\frac{1}{p}}$$

It can be written

$$\mathcal{E}_p(\Gamma_\ell) = \left( \sum_{j=1}^{\ell} \frac{\mathrm{card}\,(C_j)}{n} \mathcal{W}_p(C_j, \delta_{\gamma_j})^p \right)^{\frac{1}{p}}$$

with $C_j = \{x \in (x_i)_{i=1}^n \ : \ j = \arg\min_{j' \in \mathcal{J}} \|x - \gamma_{j'}\|\}$

# Content

# Augmented quantization

**Objective**: Find $\boldsymbol{R} = (R_1, \dots, R_\ell) \in \mathcal{R}^\ell$ and $\boldsymbol{C} = (C_1, \dots, C_\ell)$ minimising the *quantization error*

**Quantization error:** $\mathcal{E}_p(\boldsymbol{C}, \boldsymbol{R}) := \left( \sum_{j=1}^{\ell} \frac{\mathrm{card}(C_j)}{n} \mathcal{W}_p(C_j, R_j)^p \right)^{\frac{1}{p}}$

**Global error:** $\epsilon_p(\boldsymbol{C}, \boldsymbol{R}) := \mathcal{W}_p\left( \bigcup_{j=1}^{\ell} C_j, R_J \right) = \mathcal{W}_p\left( (x_i)_{i=1}^n, R_J \right)$

with $J \in \mathcal{J}$ a random variable such that $P(J = j) = \frac{\mathrm{card}(C_j)}{n}$.

**Proposition**

*The global error between a clustering $\boldsymbol{C}$ and a set of representatives $\boldsymbol{R}$ is lower that the quantization error between them:*

$$\epsilon_p(\boldsymbol{C}, \boldsymbol{R}) \leq \mathcal{E}_p(\boldsymbol{C}, \boldsymbol{R}).$$

# Lloyd's algorithm

**Algorithm** Lloyd's algorithm

**Input:** $(\gamma_1, \ldots, \gamma_\ell) \in \mathcal{X}^\ell$ , sample $(x_i)_{i=1}^n$

**while** stopping criterion not met **do**

Update clusters: $C_j \leftarrow \{x \in (x_i)_{i=1}^n \, : \, j = \arg\min_{j' \in \mathcal{J}} \|x - \gamma_{j'}\|\}$

Update representatives: $\gamma_j = \frac{1}{\mathrm{card}(C_j)} \sum_{x \in C_j} x$

**end while**

$J$ r.v. with $p_j = \mathbb{P}(J = j) = \frac{\mathrm{card}(C_j)}{n}, \quad j \in \mathcal{J}$

**Output:** $\sum_{j=1}^{\ell} p_j \delta_{\gamma_j}$

# General Lloyd's algorithm

**Algorithm** Rewritten Lloyd's algorithm

**Input:** $\boldsymbol{R} = (R_1, \ldots, R_\ell) \in \mathcal{R}^\ell$ , Sample $(x_i)_{i=1}^n$

**while** stopping criterion not met **do**
  Update clusters: $\boldsymbol{C} \leftarrow \mathit{FindC}(\boldsymbol{R})$
  Update representatives: $\boldsymbol{R} \leftarrow \mathit{FindR}(\boldsymbol{C})$
**end while**

$J$ r.v. with $p_j = \mathbb{P}(J = j) = \frac{\mathrm{card}(C_j)}{n}, \quad j \in \mathcal{J}$

**Output:** $R_J$

# Exploration problem

**What we need**:

- *FindC* providing clusters from representatives
- *FindR* providing representatives from clusters

**Problem**: Only *FindC* and *FindR* are not sufficient to be exploratory enough in the case of continuous distribution

# Augmented quantization algorithm

**Algorithm** Augmented Quantization algorithm

**Input:** $\boldsymbol{R} = (R_1, \ldots, R_\ell) \in \mathcal{R}^\ell$, samples $(x_i)_{i=1}^n$

$J \in \mathcal{J}$ r.v. with $\mathbb{P}(J = j) = \frac{1}{\ell}$
$(R^\star, C^\star, \mathcal{E}^\star) \leftarrow (\emptyset, \emptyset, +\infty)$

while stopping criterion not met do
  Update clusters: $\boldsymbol{C} \leftarrow \mathit{FindC}(\boldsymbol{R}, J)$
  Perturb clusters: $\boldsymbol{C} \leftarrow \mathit{Perturb}(\boldsymbol{C})$
  Update mixture: $\boldsymbol{R} \leftarrow \mathit{FindR}(\boldsymbol{C})$, $J$ r.v. with $\mathbb{P}(J = j) = \frac{\mathrm{card}\left(C_j\right)}{n}$, $j \in \mathcal{J}$
  Update the best configuration:
  if $\mathcal{E}_p(\boldsymbol{C}, \boldsymbol{R}) < \mathcal{E}^\star$ then $\mathcal{E}^\star \leftarrow \mathcal{E}$ , $\boldsymbol{C}^\star \leftarrow \boldsymbol{C}$ , $\boldsymbol{R}^\star \leftarrow \boldsymbol{R}$ , $J^\star \leftarrow J$
end while

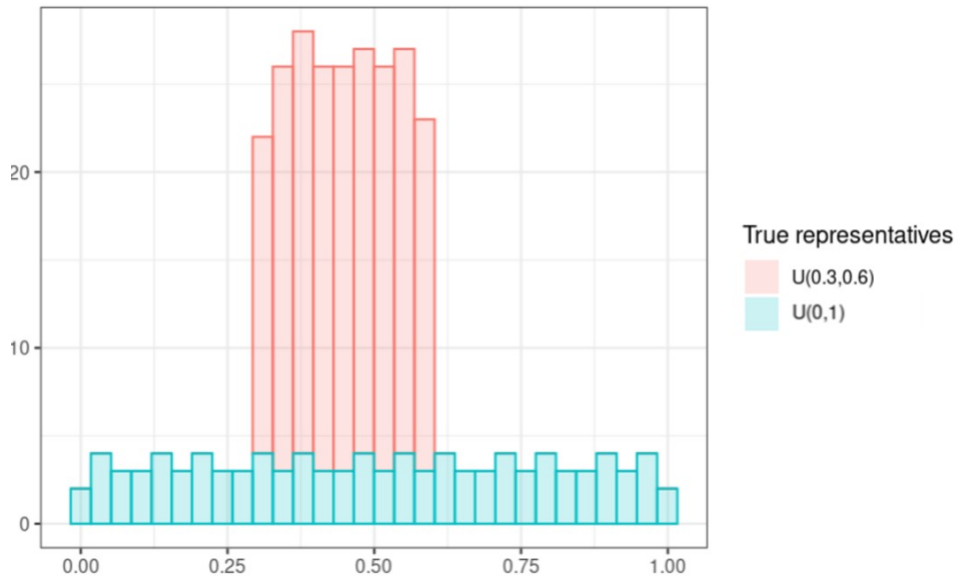
**Output:**

- the membership discrete random variable $J^\star$ with $\mathbb{P}(J^\star = j) = \frac{\mathrm{card}\left(C_j^\star\right)}{n}$, $j \in \mathcal{J}$
- the mixture $R^\star_{J^\star}$

# Illustrative sample $(x_i)_{i=1}^n$

$$\begin{cases} R_1^{\text{true}} &= R_U(0,1), \quad P(J=1) = \frac{1}{3} \\ R_2^{\text{true}} &= R_U(0.3,0.6), \quad P(J=2) = \frac{2}{3} \end{cases}$$

with $R_U(a,b)$ the measure associated to $\mathcal{U}(a,b)$




We start with $R_1 = R_U(0, 0.5)$ and $R_2 = R_U(0.5, 1)$
We investigate $\mathcal{R} = \{R_U(a,b), a \leq b\}$

# Content

# FindC

**Algorithm** *FindC*

**Input:** Sample $(x_i)_{i=1}^n$, $\boldsymbol{R} = (R_1, \ldots, R_\ell)$, $N$, $J$ r.v. $\in \mathcal{J}$

$C_j = \emptyset$, $j \in \mathcal{J}$
$(j_i)_{i=1}^N$ N independent realizations of $J$
$(y_i)_{i=1}^N$ N independent realizations, $y_i$ sampled with associated measure $R_{j_i}$
for $x \in (x_i)_{i=1}^n$ do
  $I(x) \leftarrow \underset{i=1,\ldots,N}{\arg\min} || x - y_i ||$
  $C_{j_{I(x)}} \leftarrow C_{j_{I(x)}} \cup x$
end for

**Output:** Partition $\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N) = (C_1, \ldots, C_\ell)$

# FindC

**Algorithm** *FindC*

**Input:** Sample $(x_i)_{i=1}^n$, $\boldsymbol{R} = (R_1, \ldots, R_\ell)$, $N$, $J$ r.v. $\in \mathcal{J}$

$C_j = \emptyset$, $j \in \mathcal{J}$
$(j_i)_{i=1}^N$ N independent realizations of $J$
$(y_i)_{i=1}^N$ N independent realizations, $y_i$ sampled with associated measure $R_{j_i}$
for $x \in (x_i)_{i=1}^n$ do
  $I(x) \leftarrow \underset{i=1,\ldots,N}{\arg\min} || x - y_i ||$
  $C_{j_{I(x)}} \leftarrow C_{j_{I(x)}} \cup x$
end for

**Output:** Partition $\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N) = (C_1, \ldots, C_\ell)$

# FindC

**Algorithm** *FindC*

**Input:** Sample $(x_i)_{i=1}^n$, $\boldsymbol{R} = (R_1, \ldots, R_\ell)$, $N$, $J$ r.v. $\in \mathcal{J}$

$C_j = \emptyset$, $j \in \mathcal{J}$
$(j_i)_{i=1}^N$ N independent realizations of $J$
$(y_i)_{i=1}^N$ N independent realizations, $y_i$ sampled with associated measure $R_{j_i}$
for $x \in (x_i)_{i=1}^n$ do
  $I(x) \leftarrow \underset{i=1,\ldots,N}{\arg\min} || x - y_i ||$
  $C_{j_{I(x)}} \leftarrow C_{j_{I(x)}} \cup x$
end for

**Output:** Partition $\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N) = (C_1, \ldots, C_\ell)$

# FindC

**Algorithm** *FindC*

**Input:** Sample $(x_i)_{i=1}^n$, $\boldsymbol{R} = (R_1, \ldots, R_\ell)$, $N$, $J$ r.v. $\in \mathcal{J}$

$C_j = \emptyset$, $j \in \mathcal{J}$
$(j_i)_{i=1}^N$ N independent realizations of $J$
$(y_i)_{i=1}^N$ N independent realizations, $y_i$ sampled with associated measure $R_{j_i}$
for $x \in (x_i)_{i=1}^n$ do
  $I(x) \leftarrow \underset{i=1,\ldots,N}{\arg\min} || x - y_i ||$
  $C_{j_{I(x)}} \leftarrow C_{j_{I(x)}} \cup x$
end for

**Output:** Partition $\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N) = (C_1, \ldots, C_\ell)$

# *FindC* Convergence

**Proposition**

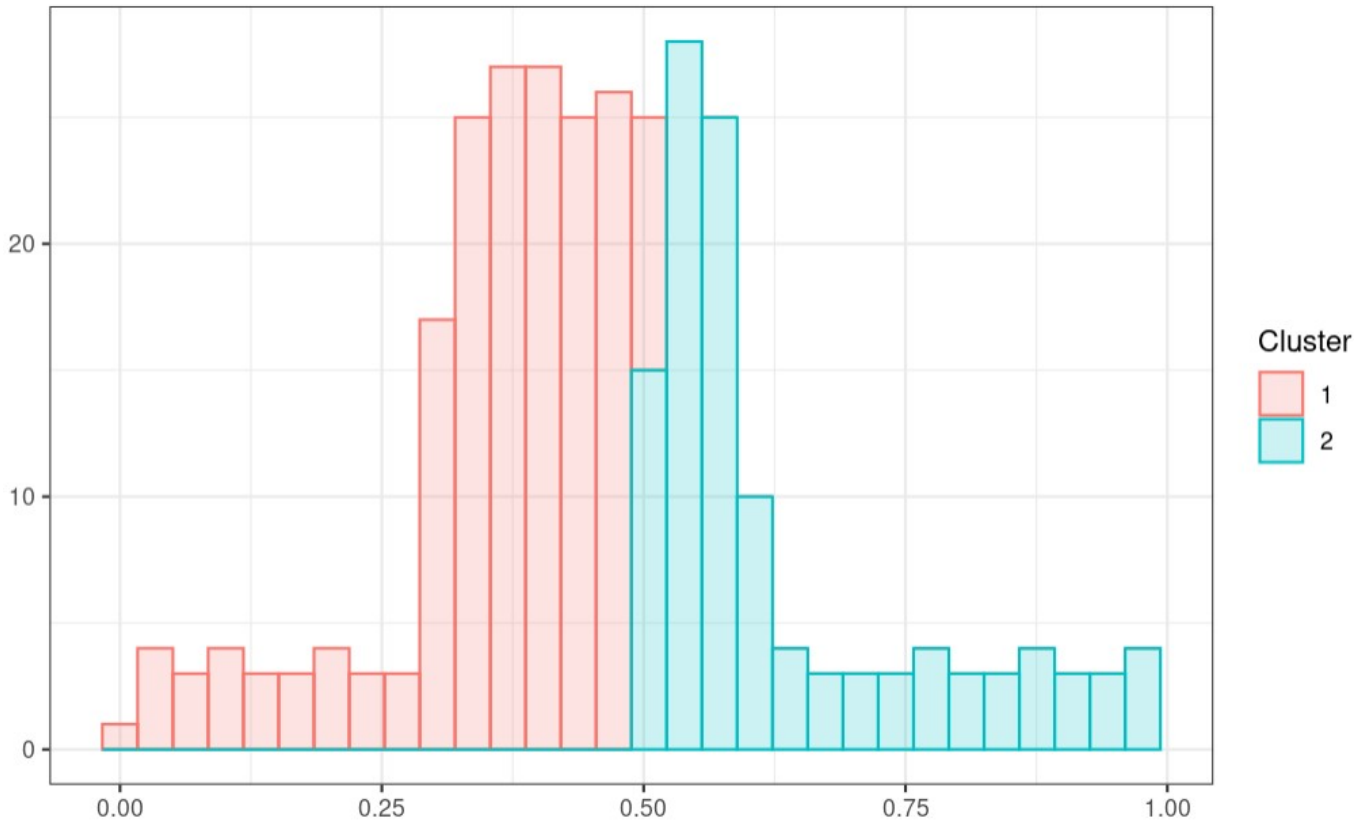
*Under general assumptions on $\mathcal{R}$, if $(X_i)_{i=1}^n$ i.i.d. sample with probability measure $R_J$ then*

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

*where $\boldsymbol{C}^\star(\boldsymbol{R}, J, n, N)$ comes from FindC*

# *FindC* illustration

Start with $R^{(1)} \sim \mathcal{U}_{[0,0.5]}$ and $R^{(2)} \sim \mathcal{U}_{[0.5,1]}$



# Content

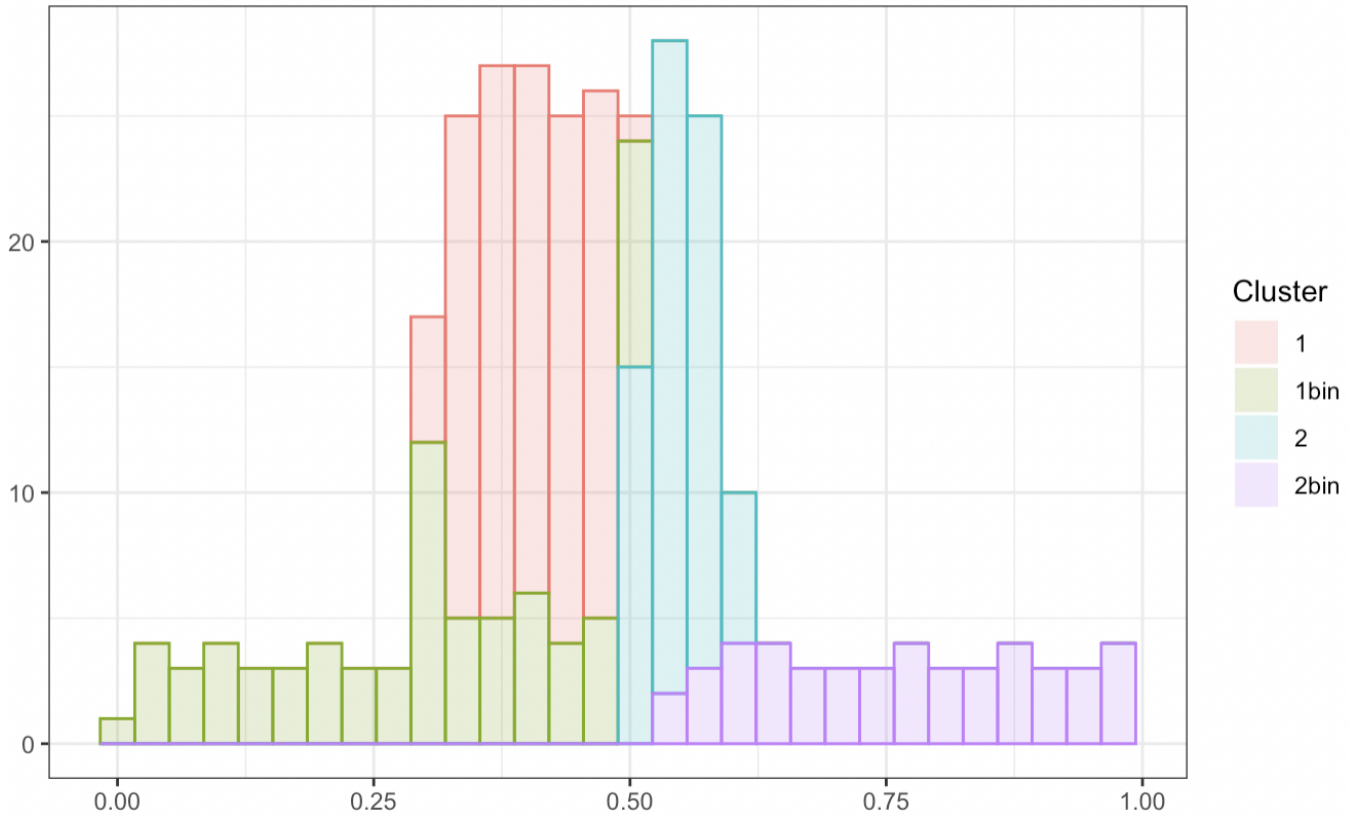
# Perturb step

**Objective:** Split the clusters by identifying their worst elements, and identify the best merge regarding the quantization error

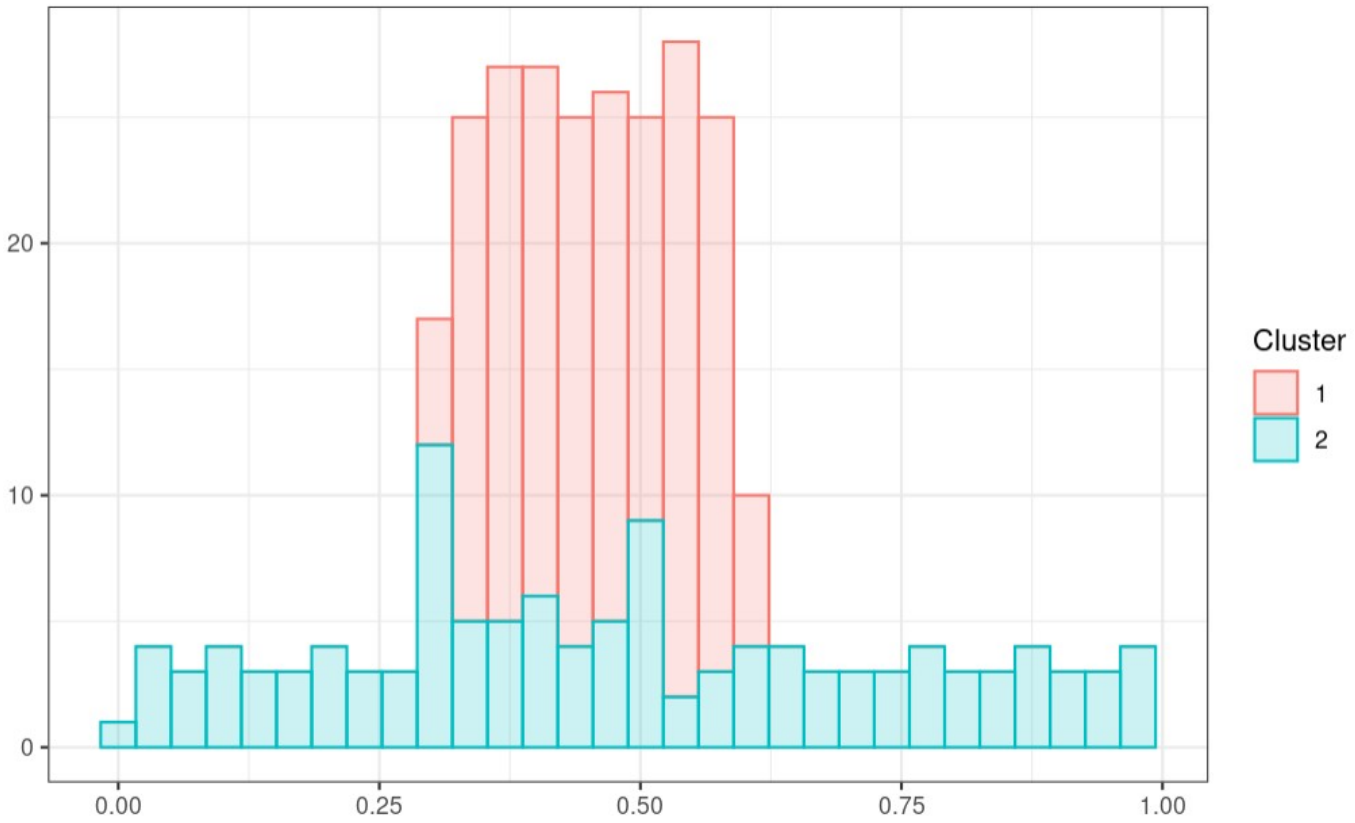
1. $\boldsymbol{C} = (C_1, \ldots, C_\ell)$
2. $\hat{C} = \text{split}(C) = (C_1, \ldots, C_\ell, C_{j_1}^{\text{bin}}, \ldots, C_{j_{\ell_{\text{bin}}}}^{\text{bin}})$
3. $\boldsymbol{C}^{\text{merge}} = \text{merge}(\hat{C}) = (C_1^{\text{merge}}, \ldots, C_\ell^{\text{merge}})$

**Important point:** The best merge can return to the clustering before the perturbation step

# Split illustration



# Merge illustration



# Content

# FindR

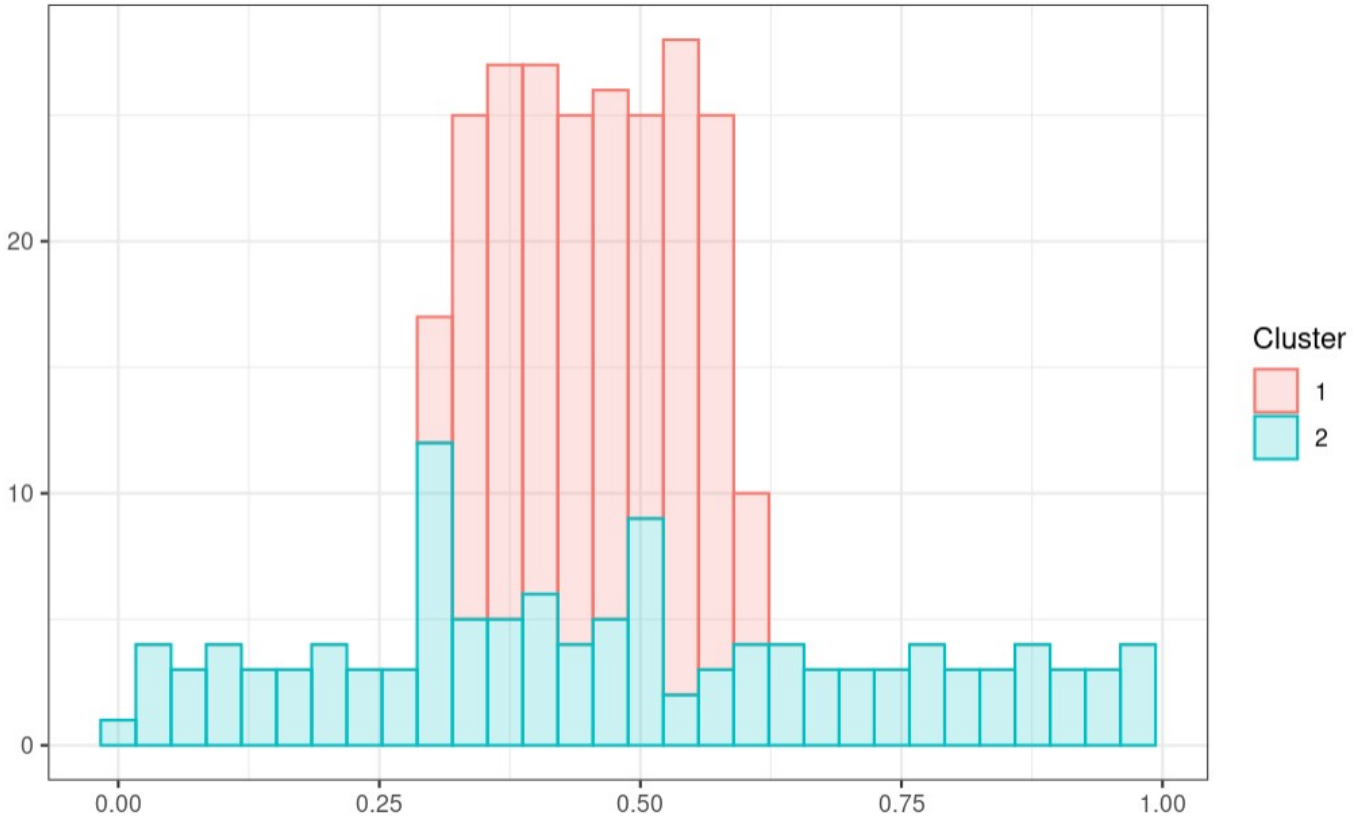
**Objective:** For a given clustering $\boldsymbol{C} = (C_1, \ldots, C_\ell)$, find its optimal representatives $\boldsymbol{R}^\star(\boldsymbol{C}) = (R_1^\star(C_1), \ldots, R_\ell^\star(C_\ell))$ with

$$R_j^\star(C_j) := \arg\min_{r \in \mathcal{R}} \mathcal{W}_p(C_j, r).$$

**General idea:** When $\mathcal{R}$ is parametric, i.e. $\mathcal{R} = \{r(\underline{\eta}), \underline{\eta} \in \mathbb{R}^q\}$, find the best parameters $\eta_k^\star$ for each marginal.

Why ? In 1D, $\mathcal{W}_p(\mu_1, \mu_2) = \left( \int_0^1 \mid F_1^{-1}(q) - F_2^{-1}(q) \mid^p dq \right)^{\frac{1}{p}}$ Panaretos et al. 2019

# *FindR* illustration



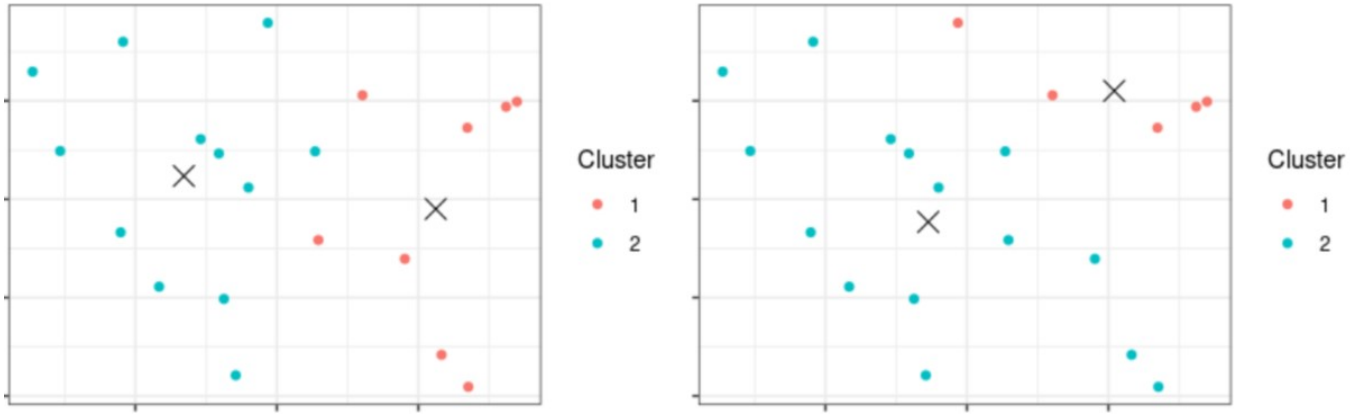

$R_1 = R_U(0.30, 0.61)$ and $R_2 = R_U(-0.02, 0.94)$

# Content

# Dirac

- *FindC* creates Voronoï cells: $x \in C_j \iff j \in \underset{j' \in \{1,\dots,\ell\}}{\arg\min} \|x - \gamma_{j'}\|$
- *FindR* identifies the centroids of the clusters $\boldsymbol{C}$

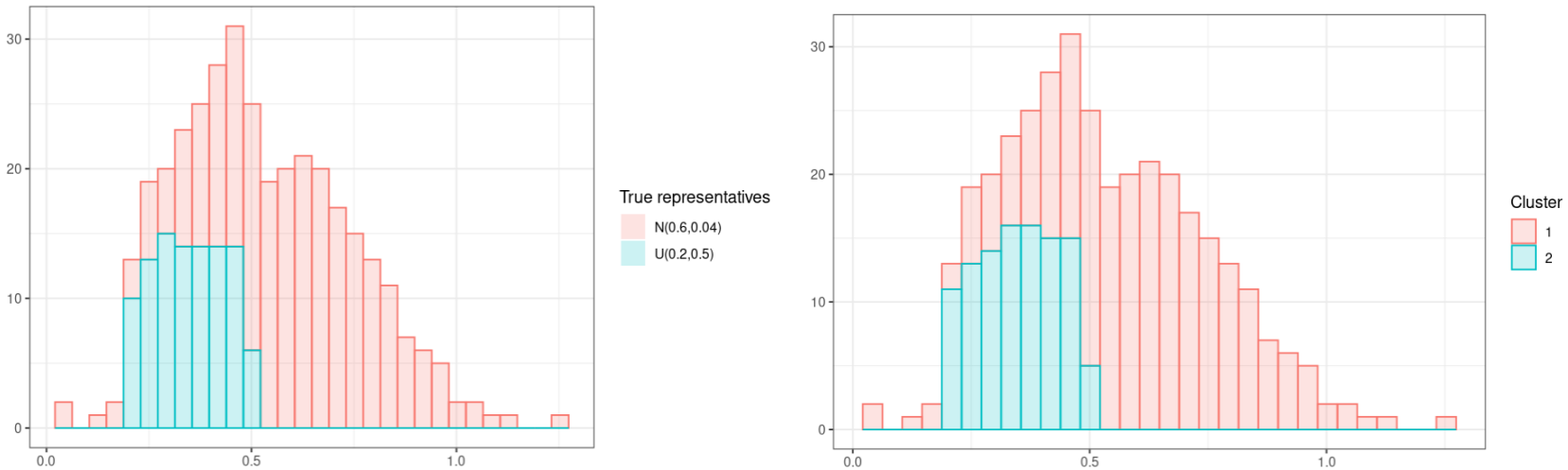
Without the clusters perturbation, AQ is equivalent to K-means




Lloyd's algorithm ($\epsilon_2(\Gamma_2) = 0.28$) Augmented quantization ($\epsilon_2(\Gamma_2) = 0.25$)

# Hybrid mixture

True representatives: $\mathcal{U}(0.2, 0.5)$ and $\mathcal{N}(0.60, 0.20^2)$.




**(a)** Disitribution of an hybrid sample

**(b)** Estimated clusters

Estimated representatives: $R_U(0.21, 0.49)$ and $R_N(0.60, 0.20^2)$.

# Content

# Possible extensions

Article submitted to Statistics and Computing
`https://arxiv.org/abs/2309.08389`

- Method to optimize the covariance structure of the representatives
- Adapt AQ with alternative metric (e.g. MMD, Smola et al. 2006)

# Bibliography I


- Dellaert, Frank (July 2003). "The Expectation Maximization Algorithm". In.
- Delyon, Bernard, Marc Lavielle, and Eric Moulines (1999). "Convergence of a stochastic approximation version of the EM algorithm". In: *Annals of statistics*, pp. 94–128.
- McLachlan, Geoffrey J., Sharon X. Lee, and Suren I. Rathnayake (2019). "Finite Mixture Models". In: *Annual Review of Statistics and Its Application* 6.1, pp. 355–378. DOI: 10.1146/annurev-statistics-031017-100325. eprint: https://doi.org/10.1146/annurev-statistics-031017-100325. URL: https://doi.org/10.1146/annurev-statistics-031017-100325.

# Bibliography II


- Nguyen, Hien D, Florence Forbes, and Geoffrey J McLachlan (2020). "Mini-batch learning of exponential family finite mixture models". In: *Statistics and Computing* 30, pp. 731–748.
- Panaretos, Victor M. and Yoav Zemel (2019). "Statistical Aspects of Wasserstein Distances". In: *Annual Review of Statistics and Its Application* 6.1, pp. 405–431. DOI: 10.1146/annurev-statistics-030718-104938.
- Smola, Alexander J, A Gretton, and K Borgwardt (2006). "Maximum mean discrepancy". In: *13th international conference, ICONIP*, pp. 3–6.
- Sridharan, Ramesh (2014). "Gaussian mixture models and the EM algorithm". In.