Warping

J-M. LOUBES

Institut de Mathématiques de Toulouse

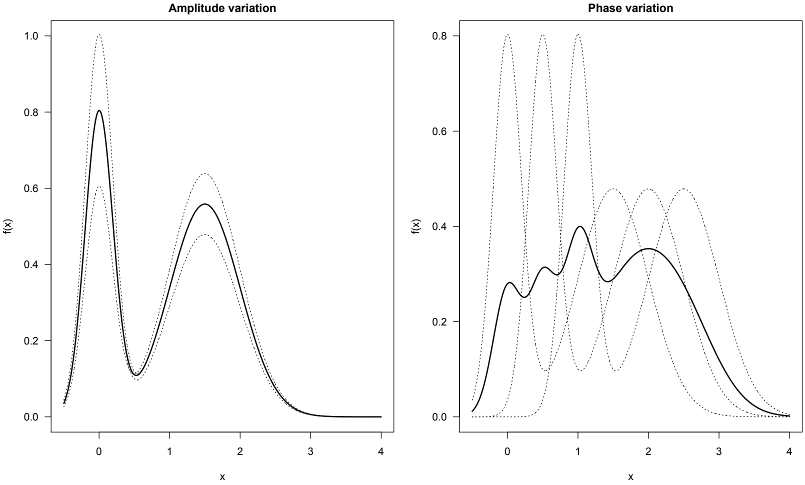
18 novembre 2011

Outline

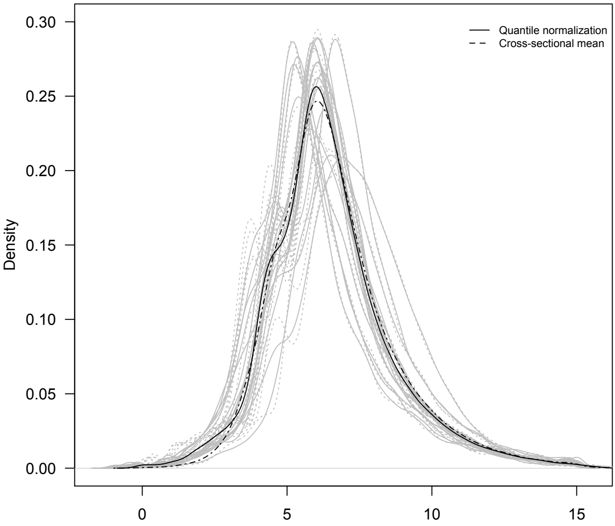
Normalization for oligonucleotide array

When running experiments that involve multiple high density oligonucleotide arrays, it is important to **remove sources of variation** between arrays of non-biological origin. **Normalization** is a process for reducing this variation. It is common to see non-linear relations between arrays and the standard normalization provided by Affymetrix **does not perform** well in these situations. Boldstad et al. 2003

Example


Amplitude variation
f(x)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
x
Phase variation
f(x)
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
x

Example


Quantile normalization
Cross-sectional mean
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15

Curve warping model

The regression model:

$$Y_{i,j} = f_j^\star(t_{ij}) + \sigma\epsilon_{i,j}, \quad i = 1,\ldots,n,\ j = 1,\ldots,J.$$

where

- $f_j^\star$ models the j^{th} signal (unknown);
- t_{ij} the observation points (known).
- $\epsilon_{i,j}$ is white noise (unknown), and σ variance (unknown)

Assumption: There exists a **common shape** of the signal $f^\star$ and **warping operators** Φ_j,

$$f_j^\star = \Phi_j f^\star, \quad j = 1,\ldots,J.$$

Aim: Estimation of the deformations and the template $f^\star$

Question : registration procedure ?

Models for curve warping

- $\Phi = \Phi_\theta$:
  parametric model for deformations =>
  Semiparametric statistics

$$\theta = (a, b, \upsilon)', \quad \Phi_\theta : f(\cdot) \to af(\cdot - b) + \upsilon$$

$$(\theta) \to \frac{1}{J}\sum_{j=1}^{J} \left\| g_j(\theta, x) - \frac{1}{J}\sum_{j'=1}^{J} g_{j'}(\theta, x) \right\|_{L^2},$$

  where $g_j(\theta, x) = \Phi_\theta \circ f_j^\star(x)$.
  M-estimators of the parameters $\hat{\theta}$
  well studied in Gamboa JML Maza (2007), Vimond (2009).
- **Non parametric framework** : Random warping process
  $h_j \sim_{i.i.d} H : \Omega \to \mathcal{C}([a, b])$
  i) $H(w, \cdot)$ is an increasing function,
  ii) $H(w, a) = a$ and $H(w, b) = b$.

$$f_j = f \circ h_j, \; j = 1, \ldots, J$$

Structural expectation

$$Y_{ij} = f_i(t_{ij}) = f \circ h_i^{-1}(t_{ij}),\ i = 1, \ldots, n,\ j = 1, \ldots, J. \qquad (1)$$

mean of the process $\phi(x) = \mathbf{E}[H(w, x)]$

- **Not identiability** $=>$ f can not be estimated hence problem $=$ definition of a mean pattern (information) that can be **recovered**
- either choosing a particular curve ... problem of arbitrary choice
- Structural expectation : takes into account the deformation

$$f_{ES} := f \circ \phi^{-1}.$$

$$f_i = f \circ h_i^{-1} = f_{ES} \circ \phi \circ h_i^{-1}$$

Mean pattern taking into account the mean deformation

Estimation of Structural expectation

Assumption : f increasing function

$$f_i = f \circ h_i^{-1} \Rightarrow f_i^{-1} = h_i \circ f^{-1} \Rightarrow \mathbf{E}(f_i^{-1}) = (\mathbf{E}(H)) \circ f^{-1}$$

$$\forall y,\ j_i(y) = \arg\min_{j \in \{1,\ldots,J\}} |Y_{ij} - y| \quad \text{and} \quad T_i(y) := t_{ij_i(y)}.$$

Empirical estimator of the inverse of the structural expectation

$$\widehat{f_{ES}^{-1}}(y) = \frac{1}{n}\sum_{i=1}^{n} T_i(y).$$

$\widehat{f_{ES}^{-1}}$: increasing step function with jumps at $K(n,J)$ points $v_1, \ldots, v_{K(n,J)}$ in $[f(a), f(b)]$, such that $f(a) = v_0 < v_1 < \ldots < v_{K(n,J)} < v_{K(n,J)+1} = f(b)$.

$$\widehat{f_{ES}^{-1}}(y) = \sum_{k=0}^{K(n,J)} u_k \mathbf{1}_{(v_k, v_{k+1})}(y)$$

Estimation of Structural expectation Dupuy JML Maza (2011)

Construction of estimator by **interpolation**:

$$\widehat{f_{ES}}(t) = \sum_{k=0}^{K(n,J)-1} \left(v_k + \frac{v_{k+1} - v_k}{u_{k+1} - u_k}(t - u_k) \right) \mathbf{1}_{[u_k, u_{k+1})}(t) + v_{K(n,J)} \mathbf{1}_{\{b\}}(t).$$

Theorem (Consistency of structural expectation estimator and warping individual function $J > \sqrt{n}$)

$$\left\| \widehat{f_{ES}} - f_{ES} \right\|_\infty \xrightarrow[n,J\to\infty]{as} 0. \forall i_0, \quad \left\| \widehat{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1} \right\|_\infty \xrightarrow[n,J\to\infty]{as} 0.$$

- Breaking monotonicity with monotonizing operator conserving the warping paths
- Observations with noise : denoising with kernel estimates

Model : Extension to points cloud

- $X_{i,j},\ i = 1, \dots, m,\ j = 1, \dots, n_i$ be a sample of m independent real valued random variables of size n_i with density function $f_i \colon \mathbb{R} \to [0, +\infty)$ and distribution function $F_i \colon \mathbb{R} \to [0, 1]$.
- Each distribution function F_i is obtained by warping a common distribution function $F \colon \mathbb{R} \to [0, 1]$ by an invertible and differentiable warping function H_i

$$F_i(t) = \Pr(X_{i,j} \leq t) = F \circ H_i^{-1}(t), \qquad i = 1, \dots, m,\ j = 1, \dots, n.$$

Model

consider the *structural expectation (SE)* of the quantile function to overcome this problem as

$$q_{SE}(\alpha) := F_{SE}^{-1}(\alpha) = \phi \circ F^{-1}(\alpha), \qquad 0 \leq \alpha \leq 1. \qquad (2)$$

Inverting equation leads to

$$q_i(\alpha) = F_i^{-1}(\alpha) = H_i \circ F^{-1}(\alpha), \qquad 0 \leq \alpha \leq 1 \qquad (3)$$

where $q_i(\alpha)$ is the population quantile function (the left continuous generalized inverse of F_i), $F_i^{-1}\colon [0,1] \to \mathbb{R}$, given by

$$q_i(\alpha) = F_i^{-1}(\alpha) = \inf\{x_{ij} \in \mathbb{R} \colon F_i(x_{ij}) \geq \alpha\}, \qquad 0 \leq \alpha \leq 1. \quad (4)$$

Hence the natural estimator of the structural expectation (2) is given by

$$\overline{q_m(\alpha)} = \frac{1}{m}\sum_{i=1}^{m} q_i(\alpha), \qquad 0 \leq \alpha \leq 1. \qquad (5)$$

A1. There exists a constant $C_1 > 0$ such that for all $(\alpha, \beta) \in [0,1]^2$, we have

$$\mathbf{E}\left[\left|H(\alpha) - \mathbf{E}H(\alpha) - \big(H(\beta) - \mathbf{E}H(\beta)\big)\right|^2\right] \leq C_1 \left|\alpha - \beta\right|^2 .$$

A2. There exists a constant $C_2 > 0$ such that, for all $(\alpha, \beta) \in [0,1]^2$, we have

$$\mathbf{E}\left[\left|F^{-1}(\alpha) - F^{-1}(\beta)\right|^2\right] \leq C_2 \left|\alpha - \beta\right|^2 .$$

Consistency

Theorem

The estimator $\overline{q_m(\alpha)}$ *is consistent is the sense that*

$$\left\|\overline{q_m(\alpha)} - \mathbf{E}\left(\overline{q_m(\alpha)}\right)\right\|_\infty = \left\|\overline{q_m(\alpha)} - q_{SE}(\alpha)\right\|_\infty \xrightarrow[m\to\infty]{a.s.} 0.$$

Moreover, under assumptions [A1] and [A2], the estimator is asymptotically Gaussian, for any $K \in \mathbb{N}$ *and fixed* $(\alpha_1, \ldots, \alpha_K) \in [0,1]^K$,

$$\sqrt{m}\begin{bmatrix} \overline{q_m(\alpha_1)} - q_{SE}(\alpha_1) \\ \vdots \\ \overline{q_m(\alpha_K)} - q_{SE}(\alpha_K) \end{bmatrix} \xrightarrow[m\to\infty]{\mathcal{D}} \mathcal{N}_K(\mathbf{0}, \Sigma)$$

where $\Sigma_{k,k'} = \vartheta\big(q(\alpha_k), q(\alpha_{k'})\big)$ for all $(\alpha_k, \alpha_{k'}) \in [0,1]^2$ with $\alpha_k < \alpha_{k'}$.

Estimation

Consider the order statistics $X_{i,1:n} \leq X_{i,2:n} \leq, \ldots, \leq X_{i,n:n}$, hence the estimation of the quantile functions, $q_i(\alpha)$, is obtained by

$$\hat{q}_{i,n}(\alpha) = \mathbb{F}_{i,n}^{-1}(\alpha) = \inf\{x_{ij} \in \mathbb{R} \colon \mathbb{F}_{i,n}(x_{ij}) \geq \alpha\}$$
$$= X_{i,j:n} \quad \text{for} \quad \frac{j-1}{n} < \alpha \leq \frac{j}{n}, \qquad j = 1, \ldots, n. \tag{6}$$

where $\mathbb{F}_{i,n}^{-1}$ is the ith empirical quantile function.

Finally, the estimator of the structural quantile is given by

$$\bar{\hat{q}}_j = \frac{1}{m}\sum_{i=1}^{m} \hat{q}_{i,j} = \frac{1}{m}\sum_{i=1}^{m} X_{i,j:n}, \qquad j = 1, \ldots, n. \tag{7}$$

Note that, this procedure corresponds to the so-called quantile normalization method proposed by Bolstad-03.

Consistency

Theorem

The quantile normalization estimator $\bar{\hat{q}}_j$ is strongly consistent

$$\bar{\hat{q}}_j \xrightarrow[m,n\to\infty]{a.s} q_{SE}(\alpha_j), \qquad j = 1, \ldots, n,$$

and under the assumptions of compactly central data, $|X_{i,j:n} - \mathbf{E}(X_{i,j:n})| \leq L < \infty$ for all i and j, and $\frac{\sqrt{m}}{n} \to 0$, it is asymptotically Gaussian. Actually, for any $K \in \mathbb{N}$ and fixed $(\alpha_1, \ldots, \alpha_K) \in [0,1]^K$,

$$\sqrt{m} \begin{bmatrix} \bar{\hat{q}}_{j_1} - q_{SE}(\alpha_1) \\ \vdots \\ \bar{\hat{q}}_{j_K} - q_{SE}(\alpha_K) \end{bmatrix} \xrightarrow[m,n\to\infty]{\mathcal{D}} \mathcal{N}_K(\mathbf{0}, \Sigma)$$

where $\Sigma_{k,k'} = \vartheta\big(q(\alpha_k), q(\alpha_{k'})\big)$ for all $(\alpha_k, \alpha_{k'}) \in [0,1]^2$ with

Asymptotic behavior of the quantile estimator, $\hat{q}_{i,n}(\alpha)$

Theorem

Assume F_i is continuously differentiable at the αth population quantile $q_i(\alpha)$ which is the unique solution of $F_i(q_i(\alpha)-) \leq \alpha \leq F_i(q_i(\alpha))$, and $f_i\big(q_i(\alpha)\big) > 0$ for a fixed $0 < \alpha < 1$. Also assume $n^{-1/2}(j/n - \alpha) = o(1)$. Then, for $i = 1, \ldots, m$, the estimator $\hat{q}_{i,n}(\alpha)$ is strongly consistent,

$$\hat{q}_{i,n}(\alpha) \xrightarrow[n\to\infty]{a.s.} q_i(\alpha)$$

$$\sqrt{n}\big(X_{i,j:n} - H_i \circ q(\alpha)\big) \xrightarrow[n\to\infty]{\mathcal{D}}$$

$$\mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{\left(f\circ H_i^{-1}\big(H_i\circ q(\alpha)\big)\cdot\big(H_i^{-1}\big)'\big(H_i\circ q(\alpha)\big)\right)^2}\right)$$

where $\big(H_i^{-1}\big)'(z) = \frac{dH_i^{-1}(z)}{dz} = \frac{1}{H_i'\circ H_i^{-1}(z)}$.

Proofs

$X_{i,j:n} \stackrel{d}{=} F_i^{-1}(U_{i,j:n})$ around the point $\mathbf{E}(U_{i,j:n}) = \alpha_j = j/(n+1)$, where $U_{i,j:n}$ denotes the jth order statistic in a sample of size n from the uniform $(0,1)$ distribution. The approximated means, variances and covariances of order statistics for $i = 1, \ldots, m$ are given by

$$\mathbf{E}(X_{i,j:n}) = q_{i,j} + \frac{\alpha_j(1-\alpha_j)}{2(n+2)} q''_{i,j} + \frac{\alpha_j(1-\alpha_j)}{(n+2)^2}\left[\frac{1}{3}\big((1-\alpha_j)-\alpha_j\big) q'''_{i,j} + \frac{1}{8}\alpha_j(1-\alpha_j) q^{(4)}_{i,j}\right] + O\left(\frac{1}{n^2}\right) \quad (8)$$

$$\mathbf{Var}(X_{i,j:n}) = \frac{\alpha_j(1-\alpha_j)}{n+2} q'^2_{i,j} + \frac{\alpha_j(1-\alpha_j)}{(n+2)^2}\left[2\big((1-\alpha_j)-\alpha_j\big) q'_{i,j} q''_{i,j} + \alpha_j(1-\alpha_j)\left(q'_{i,j} q'''_{i,j} + \frac{1}{2} q''^2_{i,j}\right)\right] + O\left(\frac{1}{n^2}\right) \quad (9)$$

Playing with the functions

One of the major issue in registration problem is to find the fitting criterion which enables to give a sense to the notion of mean of a sample of points. A natural criterion is in this framework given by the Wasserstein distance and this problem can be rewritten as finding a measure μ which minimizes

$$\mu \mapsto \frac{1}{m} \sum_{i=1}^{m} W_2^2(\mu, \mu_i), \tag{10}$$

where W_2 stands for the 2-Wasserstein distance

$$W_2^2(\mu, \mu_i) = \int |F_i^{-1}(t) - F^{-1}(t)|^2 dt.$$

Extension

for any distance d on the inverse of distribution functions, we can define a criterion to be minimized

$$F \mapsto \frac{1}{m}\sum_{i=1}^{m} d(F^{-1}, F_i^{-1}).$$

Each choice of d implies different properties for the minimizers. Recall that the choice of the L^2 loss corresponds to the Wasserstein distance between the distributions. Another choice, when dealing with warping problems, is to consider that the functional data belong to a non euclidean set, and to look for the most suitable corresponding distance. Hence, a natural framework is given by considering that the functions belong to a manifold using a manifold embedding

Extension

$\hat{d}_g$, an approximation of the geodesic distance, is provided using an Isomap-type graph approximation, following Tenenbaum2000. This gives rise to the criterion

$$F \mapsto \frac{1}{m} \sum_{i=1}^{m} \hat{d}_g(F^{-1}, F_i^{-1}).$$

Question :

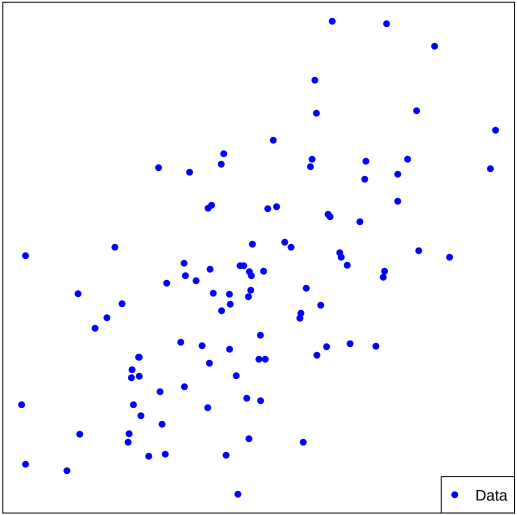
- How to choose the manifold embedding (non unique) ?
- Is there an (optimal) way to estimate the distance ?
- Notions of stability

Extension : practical implementation

$X_{i,j},\ i=1,\dots,m,\ j=1,\dots,n$ random variables. In order to mimic the geodesic distance between the inverse of the distribution functions,

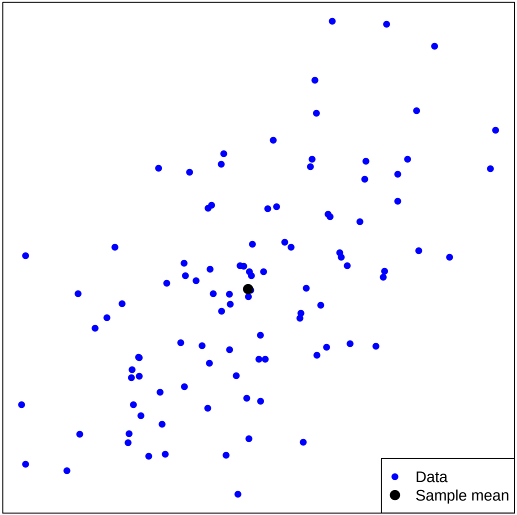
1. Estimate $F_i^{-1}(t)$, for $k-1/n < t \leq k/n$ by the corresponding order statistics $X_{i,k:n}$.
2. Sort the observations for each sample i, and denote by $X_{(i).}$ the sorted vector $X_{i,1:n},\dots,X_{i,n:n}$ and thus we obtain an array of sorted observations $(X_{(1).},\dots,X_{(m).})$.
3. Compute $\hat{d}_g$ an approximation of the geodesic distance between the vectors $X_{(i).}$
4. Hence the corresponding geodesic mean as the minimizer over all the observation vectors $x \in \{X_{(i).},\ i=1,\dots,m\}$ of the criterion

$$x \mapsto \frac{1}{m}\sum_{i=1}^{m}\hat{d}_g(x, X_{(i).}).$$



Gaussian data :

$$X_1, X_2, \ldots, X_n$$

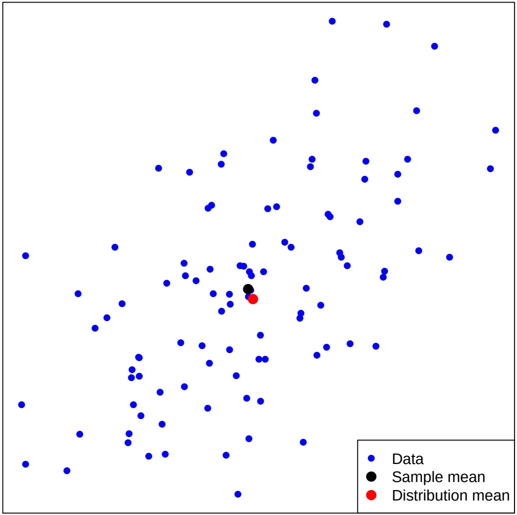


Gaussian data :

$$X_1, X_2, \ldots, X_n$$

Classical sample mean :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$



Gaussian data :

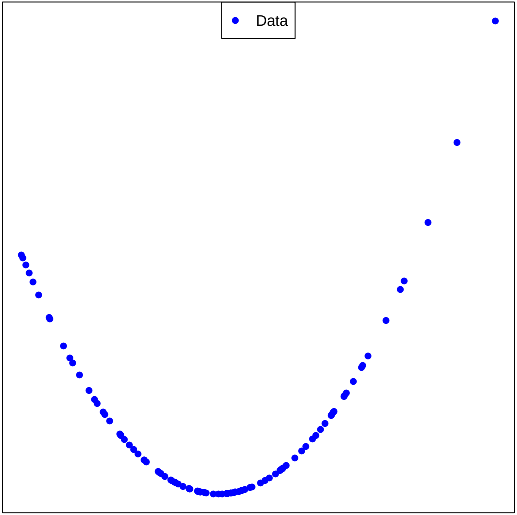
$$X_1, X_2, \ldots, X_n$$

Classical sample mean :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

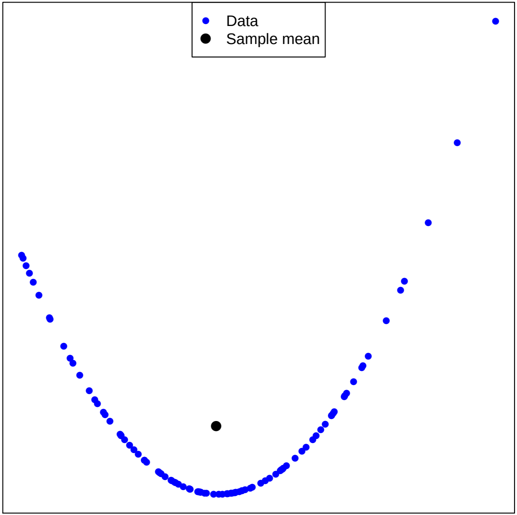
Distribution mean :

$$\bar{X} \xrightarrow[n \to +\infty]{\mathbf{P}} \mu$$



Data :

$$X_1, X_2, \ldots, X_n$$

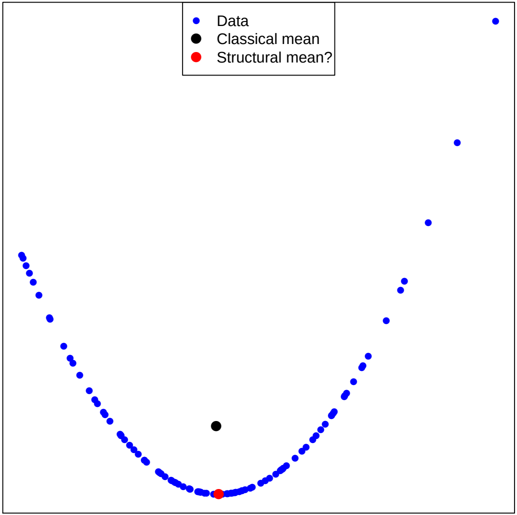


Data :

$$X_1, X_2, \ldots, X_n$$

Classical sample mean :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$



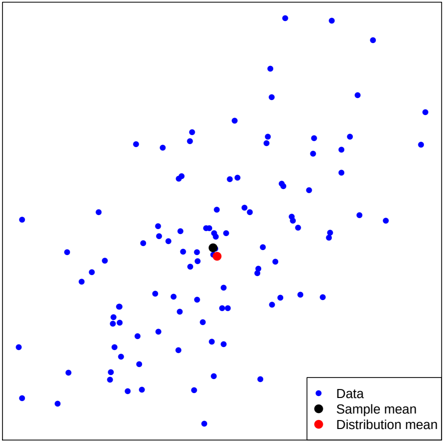
Data :

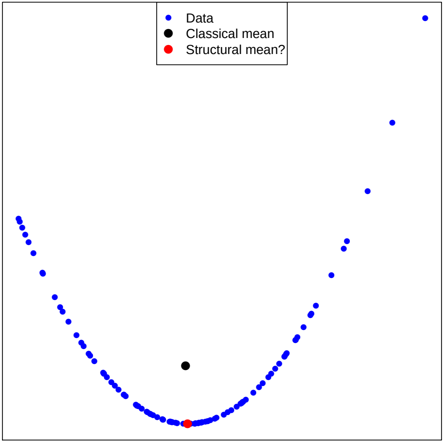
$$X_1, X_2, \ldots, X_n$$

Classical sample mean :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Structural mean ?

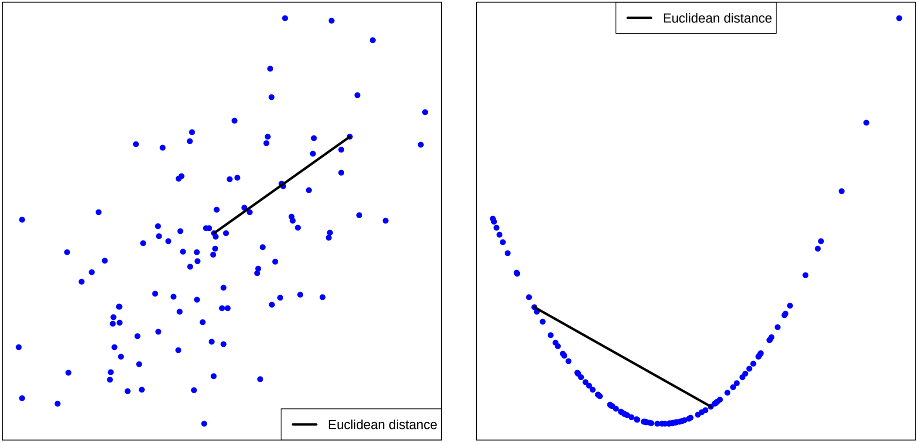

Data
Sample mean
Distribution mean


Data
Classical mean
Structural mean?

We have

$$\bar{X} = \arg\min_{a \in \mathbb{R}^2} \sum_{i=1}^{n} \mathrm{d}\left(X_i, a\right)$$

with d the Euclidean distance.


Euclidean distance
Euclidean distance

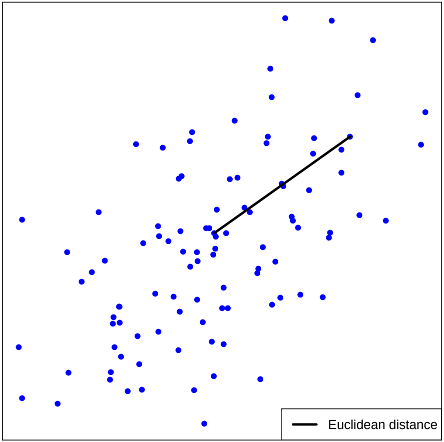
We replace

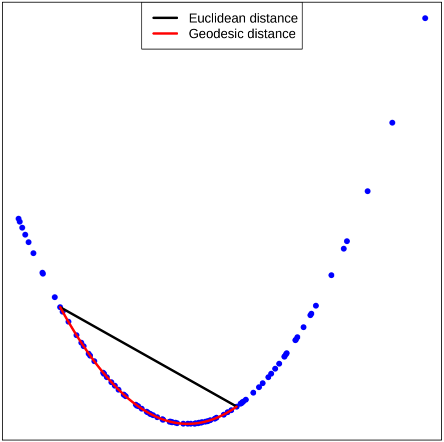
$$\arg\min_{a\in\mathbb{R}^2}\sum_{i=1}^{n}\mathrm{d}\left(X_i,a\right)$$

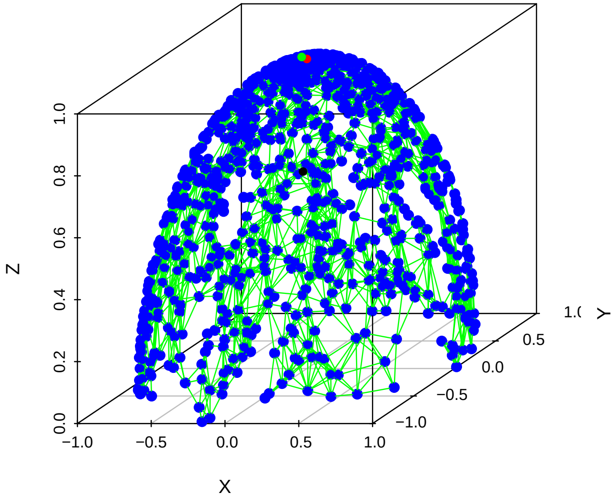
with d the Euclidean distance, by

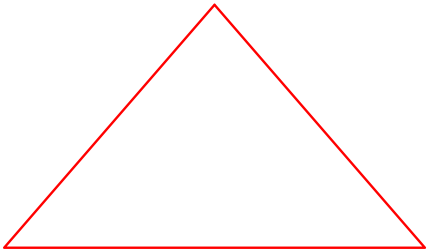
$$\arg\min_{a\in\mathcal{M}}\sum_{i=1}^{n}\delta\left(X_i,a\right)$$

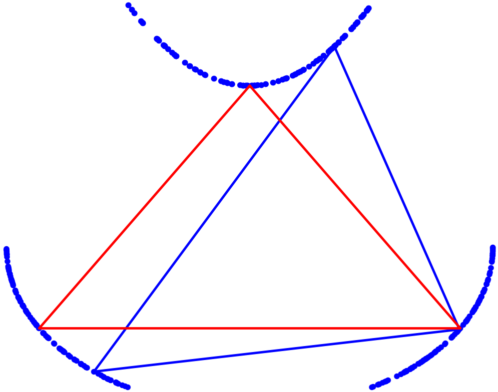
with δ the geodesic distance.

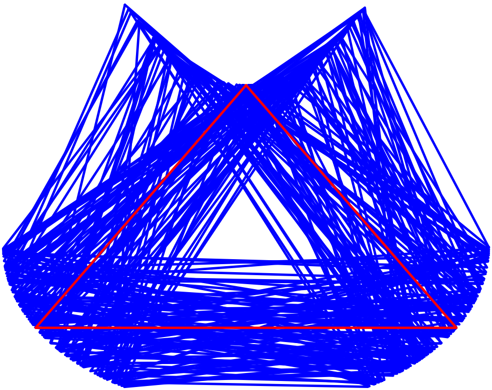

Euclidean distance

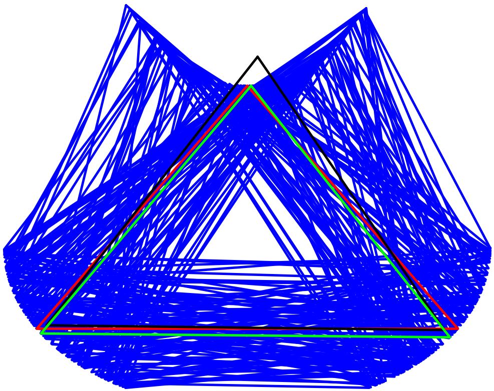

Euclidean distance
Geodesic distance

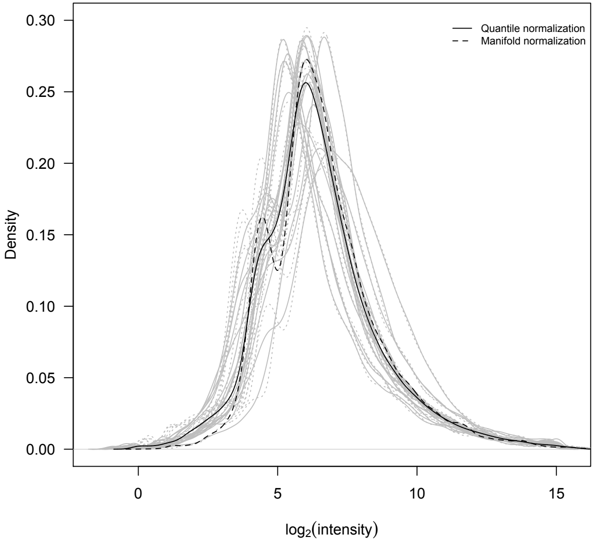

Z
X
Y
0.0
0.2
0.4
0.6
0.8
1.0
−1.0
−0.5
0.0
0.5
1.0
−1.0
−0.5
0.0
0.5
1.(










Quantile normalization
Manifold normalization
Density
log2(intensity)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15

Wasserstein Analysis

- $j = 1, \dots, J$
$$X_{ij} \sim \mu_j \quad \text{i.i.d}$$
$i = 1, \dots, n$
- μ_j comes from a family of deformations $\mu_j = T_{j\#}\mu$
- Objective : recover the unknown distribution μ and study the deformations
- Observations enable to recover the empirical distribution

$$\mu_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_{ij}}$$

Wasserstein Analysis

We say that the measure $\mu \in \mathcal{P}_2(E)$ is a Wasserstein barycenter for the measures $\mu_1, \ldots, \mu_J \in \mathcal{P}_2(E)$ endowed with weights $\lambda_1, \ldots, \lambda_J$, where $\lambda_j \geq 0, \leq j \leq J$, and $\sum_{j=1}^{J} \lambda_j = 1$, if μ minimizes

$$E(\nu) = \sum_{j=1}^{J} \lambda_j W_2^2(\nu, \mu_j).$$

We will write

$$\mu = \mathrm{Bar}((\mu_j, \lambda_j)_{1 \leq j \leq J}).$$

Wasserstein

Theorem

Let $J \geq 1$, and for every $n \geq 0$, let $\mu_j^n \in \mathcal{P}_2(\mathbb{R}^d)$, $1 \leq j \leq J$, be measures absolutely continuous w.r.t. Lebesgue measure. Let $\lambda_1, \ldots, \lambda_J$ be positive weights. Let

$$\hat{\mu}^n = Bar((\mu_j^n, \lambda_j)_{1 \leq j \leq J}).$$

Let $\mu_1, \ldots, \mu_J \in \mathcal{P}_2(\mathbb{R}^d)$ be absolutely continuous w.r.t. Lebesgue measure, and let

$$\mu^* = Bar((\mu_j, \lambda_j)_{1 \leq j \leq J}).$$

Assume that for $1 \leq j \leq J$, $W_2(\mu_j^n, \mu_j) \rightarrow 0$ for $1 \leq j \leq J$, then $W_2(\hat{\mu}^n, \mu^*) \rightarrow 0$.

Consistency of Wasserstein Barycenter

Theorem

Assume that $(T_i)_{i\in I}$ is an admissible family of deformations on a domain $\Omega \subset \mathbb{R}^n$, and let $\mu \in \mathcal{P}_2(\Omega)$, $\mu << \lambda$. Let $\mu_j = (T_j)_{\#}\mu$. The following holds :

$$IB((\mu_j, \lambda_j)_{1\leq j\leq J}) = (\sum_{j=1}^{J} \lambda_j T_j)_{\#}\mu.$$

Let γ_ε denote a $\mathcal{N}(0,\varepsilon)$ measure. Set

$$\widehat{\mu_j^n} = \mu_j^n * \gamma_{1/n}.$$

Set $\widehat{\mu_B} = \text{Bar}(\widehat{\mu_j^n}, \frac{1}{J})$. As $n \to +\infty$, we have

$$\widehat{\mu_B^n} \to \mu_B.$$

in W_2 distance

PCA Analysis

- μ_B is a suitable *mean* of the data
- Compute the distance with respect to this mean (i.e a way to center the data)
- Show on the paper ....