Estimation of deformation between distributions by minimal Wasserstein distance

Hélène Lescornel

Institut de Mathématiques de Toulouse

Colloque Jeunes Probabilistes et Statisticiens - 7/04/2014

Summary

Summary

- We observe a random variable and a deformation of this variable

$$\begin{cases} \varepsilon \\ X = \varphi(\varepsilon) \end{cases}$$

$\Rightarrow$ Random experiments with some variability : φ.

- We observe a random variable and a deformation of this variable

$$\begin{cases} \varepsilon \\ X = \varphi(\varepsilon) \end{cases}$$

$\Rightarrow$ Random experiments with some variability : φ.

- How to extract information? : estimation of the deformation, study of a mean distribution if several deformations are observed...

Warped curves

- Dynamic Time Warping. **Sakoe-Chiba-[1978]**
  Align two signals $(f(i))_{1\leqslant i\leqslant N}$ and $(g(j))_{1\leqslant j\leqslant M}$ by a time axis re normalization.
  Idea : to consider some "warping operators" between $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant M$.
  $\rightsquigarrow$ Minimize a cost

$$C(w, f, g) = \sum_{(i,j)\in w} (f(i) - g(j))^2 .$$

Warped curves

- Dynamic Time Warping. **Sakoe-Chiba-[1978]**
  Align two signals $(f(i))_{1\leqslant i\leqslant N}$ and $(g(j))_{1\leqslant j\leqslant M}$ by a time axis re normalization.
  Idea : to consider some "warping operators" between $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant M$.
  $\rightsquigarrow$ Minimize a cost

$$C(w,f,g) = \sum_{(i,j)\in w} (f(i)-g(j))^2 .$$

- Extension to warped curves in a regression scheme : **Wang-Gasser-[1999]**, **Gamboa-Loubes-Maza-[2007]**. Different cost functions.

Deformation of distributions

- Observations $X_j = \varphi_j(\varepsilon)$, $1 \leqslant j \leqslant J$ where φ_j are realizations of a random process.
  Estimation of a mean distribution using the quantile functions F_j^{-1} in **Gallòn-Loubes-Maza-[2013]**.

Deformation of distributions

- Observations $X_j = \varphi_j\left(\varepsilon\right),\ 1 \leqslant j \leqslant J$ where φ_j are realizations of a random process.
  Estimation of a mean distribution using the quantile functions F_j^{-1} in **Gallòn-Loubes-Maza-[2013]**.
- Test for a parametric relationship between two distributions in terms of quantile functions $F^{-1} = \mathcal{F}\left(G^{-1}, \theta\right)$. Test statistic based on the L^2 norm between quantile functions in **Freitag-Munk-[2005]**.

Model studied

Semi parametric framework

$$\begin{cases} \varepsilon \\ X = \varphi_{\theta^\star}(\varepsilon) \end{cases}$$

- shape of the deformation φ assumed **known**,
- deformation parameter $\theta^\star$ and template measure μ of ε to estimate.

Model studied

Semi parametric framework

$$\begin{cases} \varepsilon \\ X = \varphi_{\theta^\star}(\varepsilon) \end{cases}$$

- shape of the deformation φ assumed **known**,
- deformation parameter $\theta^\star$ and template measure μ of ε to estimate.

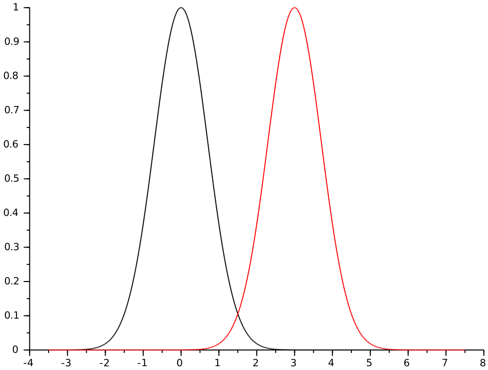
Idea: **Align** the distribution of X on the distribution of ε.
$\rightsquigarrow$ Study of $Z(\theta) = \varphi_\theta^{-1}(X) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon)$.

Parallel with warped curves

$$\varphi_\theta(t) = t + \theta$$

Parallel with warped curves

$$\varphi_\theta(t) = t + \theta$$



We have $Z(\theta^\star) = \varepsilon$.
$\rightarrow$ Align the distribution of $Z(\theta)$ on the distribution of ε by varying θ. Minimization of a D.T.W. criterion.

We have $Z(\theta^\star) = \varepsilon$.
$\rightarrow$ Align the distribution of $Z(\theta)$ on the distribution of ε by varying θ. Minimization of a D.T.W. criterion.

$\rightarrow$ Which cost function? Requires a distance between probabilities. Utilization of the **Wasserstein distance** related to problems of mass transport.

Summary

Presentation of the model

Observations :

$$\begin{cases} \varepsilon_{i1} & 1 \leqslant i \leqslant n \\ X_i = \varphi_{\theta^\star}(\varepsilon_{i2}) & 1 \leqslant i \leqslant n \end{cases} \tag{1}$$

- structure : $(\varepsilon_{ij})_{\substack{1 \leqslant i \leqslant n \\ j=1,2}}$ i.i.d. following the law μ,
- deformation parameter : $\theta^\star$ where $\theta^\star \in \Theta \subset \mathbb{R}^d$,
- deformation function : $\varphi_\theta :]a;b[\to]c;d[$ invertible for $\theta \in \Theta$,
- Empirical distribution of the i.i.d. sample $(\varepsilon_{i1})_{1 \leqslant i \leqslant n}$: μ_1^n .

For $\theta \in \Theta$ we define

$$Z_i(\theta) = \varphi_\theta^{-1}(X_i) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2})$$

- Distribution of $Z_1(\theta)$: $\mu_\star(\theta) = \mu \circ \varphi_{\theta^\star}^{-1} \circ \varphi_\theta$,
- Empirical distribution of the i.i.d. sample $(Z_1(\theta), \ldots, Z_n(\theta))$: $\mu_\star^n(\theta) = \frac{1}{n}\sum_{i=1}^n 1_{Z_i(\theta)}$.

For $\theta \in \Theta$ we define

$$Z_i(\theta) = \varphi_\theta^{-1}(X_i) = \varphi_\theta^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2})$$

- Distribution of $Z_1(\theta)$: $\mu_\star(\theta) = \mu \circ \varphi_{\theta^\star}^{-1} \circ \varphi_\theta$,
- Empirical distribution of the i.i.d. sample $(Z_1(\theta), \ldots, Z_n(\theta))$: $\mu_\star^n(\theta) = \frac{1}{n}\sum_{i=1}^n 1_{Z_i(\theta)}$.

$\Rightarrow$ Recover $\theta^\star$ by minimizing the energy needed to align the distribution $\mu_\star(\theta)$ on μ :$\mu_\star(\theta^\star) = \mu$.
Wasserstein distance to quantify the alignment.

Wasserstein distance

Set $\mathcal{W}_2(\mathbb{R}) = \left\{P \text{ probability on } \mathbb{R}, \int_{\mathbb{R}} x^2 dP < \infty\right\}$.
For $P, Q \in \mathcal{W}_2(\mathbb{R})$ with respective distribution function F and G, their Wasserstein distance is

$$W_2^2(P, Q) = \int_0^1 \left(F^{-1}(t) - G^{-1}(t)\right)^2 dt. \tag{2}$$

Wasserstein distance

Set $\mathcal{W}_2(\mathbb{R}) = \left\{ P \text{ probability on } \mathbb{R}, \int_{\mathbb{R}} x^2 dP < \infty \right\}$.
For $P, Q \in \mathcal{W}_2(\mathbb{R})$ with respective distribution function F and G, their Wasserstein distance is

$$W_2^2(P, Q) = \int_0^1 \left(F^{-1}(t) - G^{-1}(t) \right)^2 dt. \tag{2}$$

If P and Q are defined on more general metric space (S, d) with a moment of order 2 :

$$W_2^2(P, Q) = \inf_{X \sim P, Y \sim Q} \mathbb{E}\left[d(X, Y)^2 \right]$$

Wasserstein distance

Set $\mathcal{W}_2(\mathbb{R}) = \left\{ P \text{ probability on } \mathbb{R}, \int_{\mathbb{R}} x^2 dP < \infty \right\}.$
For $P, Q \in \mathcal{W}_2(\mathbb{R})$ with respective distribution function F and G, their Wasserstein distance is

$$W_2^2(P,Q) = \int_0^1 \left(F^{-1}(t) - G^{-1}(t) \right)^2 dt. \tag{2}$$

If P and Q are defined on more general metric space (S,d) with a moment of order 2 :

$$W_2^2(P,Q) = \inf_{X \sim P, Y \sim Q} \mathbb{E}\left[d(X,Y)^2 \right]$$

- Set $X_n \sim P_n, X \sim P$.
  Then $W_2(P_n, P) \to 0 \iff \begin{cases} X_n \rightharpoonup X \\ \mathbb{E}\left[X_n^2\right] \to \mathbb{E}\left[X^2\right] \end{cases}$

The estimators

To quantify the alignment of the measures, consider the criterion

$$M(\theta) = W_2^2(\mu_\star(\theta), \mu) \quad (3)$$

The estimators

To quantify the alignment of the measures, consider the criterion

$$M(\theta) = W_2^2(\mu_\star(\theta), \mu) \tag{3}$$

Characterisation of $\theta^\star : \min_\Theta M = 0 = M(\theta^\star)$.

The estimators

To quantify the alignment of the measures, consider the criterion

$$M(\theta) = W_2^2(\mu_\star(\theta), \mu) \tag{3}$$

Characterisation of $\theta^\star$: $\min_\Theta M = 0 = M(\theta^\star)$.
$\rightarrow$ Empirical version :

$$M_n(\theta) = W_2^2(\mu_\star^n(\theta), \mu_1^n)$$

with the order statistics

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[Z_{(i)}(\theta) - \varepsilon_{(i)1} \right]^2 . \tag{4}$$

with the order statistics

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[Z_{(i)}(\theta) - \varepsilon_{(i)1} \right]^2 . \qquad (4)$$

Leads to the M-estimator for the deformation parameters

Estimator of $\theta^\star$

$$\widehat{\theta}^n \in argmin_{\theta \in \Theta} M_n(\theta) . \qquad (5)$$

with the order statistics

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[Z_{(i)}(\theta) - \varepsilon_{(i)1}\right]^2. \quad (4)$$

Leads to the M-estimator for the deformation parameters

Estimator of $\theta^\star$

$$\widehat{\theta}^n \in argmin_{\theta\in\Theta} M_n(\theta). \quad (5)$$

We have $\varphi_{\widehat{\theta}^n}^{-1}(X_i) = \varphi_{\widehat{\theta}^n}^{-1} \circ \varphi_{\theta^\star}(\varepsilon_{i2}) \approx \varepsilon_{i2}$, following the unknown law μ.

with the order statistics

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[Z_{(i)}(\theta) - \varepsilon_{(i)1} \right]^2 . \tag{4}$$

Leads to the M-estimator for the deformation parameters

Estimator of $\theta^\star$

$$\widehat{\theta}^n \in argmin_{\theta \in \Theta} M_n(\theta) . \tag{5}$$

We have $\varphi^{-1}_{\widehat{\theta}^n}(X_i) = \varphi^{-1}_{\widehat{\theta}^n} \circ \varphi_{\theta^\star}(\varepsilon_{i2}) \approx \varepsilon_{i2}$, following the unknown law μ.

- Plugg-in estimator of μ

$$\widehat{\mu}^n = \frac{1}{n} \sum_{i=1}^{n} 1_{\varphi^{-1}_{\widehat{\theta}^n}(X_i)} \tag{6}$$

Idea : "increase" the size of $(\varepsilon_{i1})_{1 \leqslant i \leqslant n} \rightsquigarrow \widetilde{\mu}^n = \frac{1}{2}\left(\widehat{\mu}^n + \mu_1^n\right)$

Summary

Principle of M-estimation

→ Convergence criterion for estimators defined as minimizers of a random functional M_n on a set Θ.

$$
\begin{array}{ccc}
M_n(\theta) & \xrightarrow{n\to\infty} & M(\theta) \\
\min \uparrow & & \min \uparrow \\
\widehat{\theta}^n & \overset{?}{\to} & \theta^\star
\end{array}
$$

Principle of M-estimation

$\rightarrow$ Convergence criterion for estimators defined as minimizers of a random functional M_n on a set Θ.

$$\begin{array}{ccc} M_n(\theta) & \xrightarrow{n\to\infty} & M(\theta) \\ \min\uparrow & & \min\uparrow \\ \widehat{\theta}^n & \overset{?}{\rightarrow} & \theta^\star \end{array}$$

<u>Consistency criterion</u>

- If M is a deterministic function,
- $sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{n\to\infty} 0$ in probability,
- $\forall\delta > 0 \quad \inf_{\theta\in\Theta\cap B(\theta^\star,\delta)^c} M(\theta) > M(\theta^\star)$

then $\widehat{\theta}^n \xrightarrow{n\to\infty} \theta^\star$ in probability.

Assumptions

- Laws considered are defined on subsets of $\mathbb{R}$ and $\forall\theta \in \Theta$, $\mu_\star(\theta) \in \mathcal{W}_2(\mathbb{R})$.
  $\Rightarrow$ Computation of the Wasserstein distance.

Assumptions

- Laws considered are defined on subsets of $\mathbb{R}$ and $\forall\theta \in \Theta$, $\mu_{\star}(\theta) \in \mathcal{W}_2(\mathbb{R})$.
  $\Rightarrow$ Computation of the Wasserstein distance.
- Regularity C^1 of $\varphi_{\theta}^{-1}(x)$ with respect to $\theta \in \Theta$.

Assumptions

- Laws considered are defined on subsets of $\mathbb{R}$ and $\forall\theta \in \Theta$, $\mu_\star(\theta) \in \mathcal{W}_2(\mathbb{R})$.
  $\Rightarrow$ Computation of the Wasserstein distance.
- Regularity C^1 of $\varphi_\theta^{-1}(x)$ with respect to $\theta \in \Theta$. The family $\left\{\partial\varphi_\theta^{-1}(\,\cdot\,)\right\}_{\theta\in\Theta}$ has an envelop in $L^2(X)$:

$$\sup_{\theta\in\Theta} \left\|\partial\varphi_\theta^{-1}(x)\right\| \leqslant H(x) \text{ with } H \in L^2(X)$$

  $\Rightarrow$ Control the distance between $\mu_\star\left(\theta^1\right)$ and $\mu_\star\left(\theta^2\right)$ for $\theta^1, \theta^2 \in \Theta$.

Assumptions

- Laws considered are defined on subsets of $\mathbb{R}$ and $\forall \theta \in \Theta$, $\mu_\star(\theta) \in \mathcal{W}_2(\mathbb{R})$.
  $\Rightarrow$ Computation of the Wasserstein distance.
- Regularity C^1 of $\varphi_\theta^{-1}(x)$ with respect to $\theta \in \Theta$. The family $\left\{\partial\varphi_\theta^{-1}(\,\cdot\,)\right\}_{\theta\in\Theta}$ has an envelop in $L^2(X)$:

$$\sup_{\theta\in\Theta} \left\|\partial\varphi_\theta^{-1}(x)\right\| \leqslant H(x) \text{ with } H \in L^2(X)$$

  $\Rightarrow$ Control the distance between $\mu_\star\left(\theta^1\right)$ and $\mu_\star\left(\theta^2\right)$ for $\theta^1, \theta^2 \in \Theta$.
- Θ compact and convex subset of $\mathbb{R}^d$.
  $\Rightarrow$ Uniform convergence and Taylor expansion.

Assumptions

- Laws considered are defined on subsets of $\mathbb{R}$ and $\forall\theta \in \Theta$, $\mu_\star(\theta) \in \mathcal{W}_2(\mathbb{R})$.
  $\Rightarrow$ Computation of the Wasserstein distance.
- Regularity C^1 of $\varphi_\theta^{-1}(x)$ with respect to $\theta \in \Theta$. The family $\left\{\partial\varphi_\theta^{-1}(\,\cdot\,)\right\}_{\theta\in\Theta}$ has an envelop in $L^2(X)$:

$$\sup_{\theta\in\Theta} \left\|\partial\varphi_\theta^{-1}(x)\right\| \leqslant H(x) \text{ with } H \in L^2(X)$$

  $\Rightarrow$ Control the distance between $\mu_\star\left(\theta^1\right)$ and $\mu_\star\left(\theta^2\right)$ for $\theta^1, \theta^2 \in \Theta$.
- Θ compact and convex subset of $\mathbb{R}^d$.
  $\Rightarrow$ Uniform convergence and Taylor expansion.
- Identifiability condition : for all $\theta \neq \theta^\star$, $\varphi_\theta^{-1} \circ \varphi_{\theta^\star} \neq Id$ on a set of positive μ-measure.
  $\Rightarrow$ Uniqueness of the minimizer of the function M.

Consistency results

Deformation estimator

$$\widehat{\theta}^n \in argmin_{\theta\in\Theta} M_n\left(\theta\right) :$$

Theorem

Under previous assumptions $\widehat{\theta}^n$ converges in probability to $\theta^\star$.

Measure estimator

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\varphi^{-1}_{\widehat{\theta}^n}(X_i)}$$

Theorem

Under previous assumptions

$$W_2\left(\widehat{\mu}_n, \mu\right) \xrightarrow{n\to\infty} 0 \text{ in probability.}$$

Summary

Assumptions

In addition to the previous assumptions, we assume

- More regularity : φ^{-1} is C^2 with respect to its two variables (θ, x).

Assumptions

In addition to the previous assumptions, we assume

- More regularity : φ^{-1} is C^2 with respect to its two variables (θ, x).
- The distribution of X has a compact support with distribution function $F_\star$ C^1. We assume $F'_\star := f_\star > 0$ on its support.

$\Rightarrow$ The distribution function F associated with the law μ (law of ε) has a compact support and is C^1 with $F' = f > 0$.

Result

Set $\Phi = \int_0^1 \partial\varphi_{\theta^\star}^{-1}\left(F_\star^{-1}(t)\right)^2 dt \in \mathbb{R}^{d\times d}$.

Theorem

Under previous assumptions and if Φ is invertible, then

$$\sqrt{n}\left(\widehat{\theta}^n - \theta^\star\right) \rightharpoonup \Phi^{-1}\int_0^1 \frac{\partial\varphi_{\theta^\star}^{-1}\left(F_\star^{-1}(t)\right)}{f(F^{-1}(t))}\left[\mathbb{G}_2(t) - \mathbb{G}_1(t)\right] dt$$

where $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent standard Brownian bridges.

Idea of proof

$\rightarrow$ Remains to study $\Psi\left(F^n, F_\star^n\right)$ where F^n (resp. $F_\star^n$) is the empirical distribution function associated with the sample $(\varepsilon_{i1})_{1\leqslant i\leqslant n}$ (resp. $(X_i)_{1\leqslant i\leqslant n}$).

Idea of proof

$\rightarrow$ Remains to study $\Psi\left(F^{n}, F_{\star}^{n}\right)$ where F^{n} (resp. $F_{\star}^{n}$) is the empirical distribution function associated with the sample $(\varepsilon_{i1})_{1\leqslant i\leqslant n}$ (resp. $(X_i)_{1\leqslant i\leqslant n}$).
Convergence of the empirical distribution functions :

Theorem (Donsker)

If $Y_1, \ldots, Y_n$ are i.i.d. random variables with distribution function F and empirical distribution function F_n, the sequence $\sqrt{n}\,(F_n - F)$ converges in law in $\mathbb{S}$, the space of function cadlag on $\bar{\mathbb{R}}$ endowed with the norm $\|\cdot\|_{\infty}$ to $\mathbb{G}\circ F$ where $\mathbb{G}$ is a standard Brownian bridge.

Idea of proof

$\rightarrow$ Remains to study $\Psi(F^n, F^n_\star)$ where F^n (resp. $F^n_\star$) is the empirical distribution function associated with the sample $(\varepsilon_{i1})_{1\leqslant i\leqslant n}$ (resp. $(X_i)_{1\leqslant i\leqslant n}$).
Convergence of the empirical distribution functions :

Theorem (Donsker)

If $Y_1, \ldots, Y_n$ are i.i.d. random variables with distribution function F and empirical distribution function F_n, the sequence $\sqrt{n}(F_n - F)$ converges in law in $\mathbb{S}$, the space of function cadlag on $\bar{\mathbb{R}}$ endowed with the norm $\|\cdot\|_\infty$ to $\mathbb{G}\circ F$ where $\mathbb{G}$ is a standard Brownian bridge.

$\Rightarrow$ Application of a **Delta-method**.

Summary

Examples

- Example 1 : Translation model

$$\varphi_\theta(x) = x + \theta$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$, and $\Theta \subset \mathbb{R}$ compact interval.

Examples

- Example 1 : Translation model

$$\varphi_\theta(x) = x + \theta$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$, and $\Theta \subset \mathbb{R}$ compact interval.

$$\begin{cases} \varepsilon_{i1} \\ X_i = \theta^\star + \varepsilon_{i2} \end{cases} \quad 1 \leqslant i \leqslant n$$

$$\widehat{\theta}^n = \frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \varepsilon_{i1} = \theta^\star - \left[\frac{1}{n}\sum_{i=1}^n \varepsilon_{i2} - \frac{1}{n}\sum_{i=1}^n \varepsilon_{i1}\right]$$

- Example 2 : Logit model

$$\varphi_\theta(x) = \frac{1}{1+\exp(\theta x)}$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$ and Θ compact interval of $]-\infty; 0[$.

- Example 2 : Logit model

$$\varphi_\theta(x) = \frac{1}{1+\exp(\theta x)}$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$ and Θ compact interval of $]-\infty;0[$.

$$\left\{ \begin{array}{l} \varepsilon_{i1} \\ X_i = \frac{1}{1+\theta^\star \varepsilon_{i2}} \end{array} \right. \quad 1 \leqslant i \leqslant n$$

$$\widehat{\theta}^n = \frac{\sum_{i=1}^n \ln\left(\frac{1-X_{(i)}}{X_{(i)}}\right)^2}{\sum_{i=1}^n \ln\left(\frac{1-X_{(i)}}{X_{(i)}}\right)\varepsilon_{(i)1}} = \theta^\star \frac{\sum_{i=1}^n \varepsilon_{(i)2}^2}{\sum_{i=1}^n \varepsilon_{(i)2}\varepsilon_{(i)1}}$$

- Example 3: Location/scale model

$$\varphi_\theta(x) = \theta_2 x + \theta_1$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$ and Θ compact in $\mathbb{R} \times]0; +\infty[$.

- Example 3: Location/scale model

$$\varphi_\theta(x) = \theta_2 x + \theta_1$$

$\Rightarrow \mu \in \mathcal{W}_2(\mathbb{R})$ and Θ compact in $\mathbb{R}\times]0;+\infty[$.

$\rightarrow$ Scale model

$$\begin{cases} \varepsilon_{i1} \\ X_i = \theta^\star \varepsilon_{i2} \end{cases} \quad 1 \leqslant i \leqslant n$$

$$\widehat{\theta}^n = \frac{\sum_{i=1}^n X_{(i)}^2}{\sum_{i=1}^n X_{(i)} \varepsilon_{(i)1}} = \theta^\star \frac{\sum_{i=1}^n \varepsilon_{(i)2}^2}{\sum_{i=1}^n \varepsilon_{(i)2} \varepsilon_{(i)1}}$$

Bibliography

Gamboa-Loubes-Maza-[2007] : F. Gamboa, J.-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. Electron. J. Stat., 1 :616-640, 2007.

Gallòn-Loubes-Maza-[2011] : S. Gallòn, J-M Loubes, E. Maza,Statistical Properties of the Quantile Normalization Method for Density Curve Alignment. Technical report, May 2011.

Vimond-[2010] : M. Vimond, Efficient estimation for a subclass of shape invariant models. Ann. Statist., 38(3):1885 1912, 2010.

Wang-Gasser-[1999] : K. Wang and T. Gasser. Synchronizing sample curves nonparametrically. Ann. Statist., 27(2) :439-460, 1999.