Central Limit Theorem and bootstrap procedure for Wasserstein's barycenter variations and application to structural relationships between distributions

Eustasio del Barrio, Hélène Lescornel, Jean-Michel Loubes

Central Limit Theorem and bootstrap procedure for Wasserstein's barycenter variations and application to structural relationships between distributions

Del Barrio Eustasio, and Lescornel Hélène, and Loubes Jean-Michel

Address of the First and Second authors
IMT
118 route de Narbonne
31000 Toulouse
FRANCE
e-mail: tasio@eio.uva.es

loubes@math.univ-toulouse.fr

Abstract: Wasserstein barycenters and variance-like criterion using Wasserstein distance are used in many problems to analyze the homogeneity of collections of distributions and structural relationships between the observations. We propose the estimation of the quantiles of the empirical process of the Wasserstein's variation using a bootstrap procedure. Then we use these results for statistical inference on a distribution registration model for general deformation functions. The tests are based on the variance of the distributions with respect to their Wasserstein's barycenters for which we prove a central limit theorem.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.

1. Introduction

Analyzing the variability of large data sets is a difficult task when the information conveyed by the observations possesses an inner geometry far from the Euclidean one. Indeed, deformations on the data such as translations, scale location models for instance or more general warping procedures prevent the use of the usual methods in statistics. Looking for a way to measure structural relationships between data is of high importance. This kind of issues arises when considering the estimation of distribution functions observed with deformations. This situation occurs often in biology, for example when considering gene expression. However, when dealing with the registration of warped distributions, the literature is scarce. We mention here the method provided for biological computational issues known as quantile normalization in Gallón, Loubes and Maza (2013) and references therein. Very recently using optimal transport methodologies, comparisons of distributions have been studied using a notion of Fréchet mean for distributions, see for instance in Agueh and Carlier (2011) or a notion of depth as in Chernozhukov et al. (2014). In Czado and Munk (1998) and Munk and Czado (1998), a pioneer work in this framework enables to compare distributions in a scale location deformation model.

Actually assume we observe J samples of n i.i.d random variables $X_{i,j}$ with distribution μ_j. We define a notion of variation for such collection of distributions using the Wasserstein distance W_2 which generalizes the notion of empirical variance. For this we consider

$$V(\mu_1, \dots, \mu_J) = \inf_{\eta \in \mathcal{W}_2(\mathbb{R}^d)} \left(\frac{1}{J} \sum_{j=1}^{J} W_2^2(\mu_j, \eta) \right)^{1/2}.$$

We show how this quantity can be bootstrapped in order to build confidence tests with such criterion.

A good model for application is given when the observations come from a deformation model. More precisely it is the case when they can be written as

$$X_{i,j} = \left(\varphi_j^\star\right)^{-1}(\varepsilon_{i,j}),$$

for $j = 1, \dots, J$ where $(\varepsilon_{i,j})$ defined for all $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant J$ are i.i.d. random variables with unknown distribution μ and for deformation functions $\varphi_j^\star$. This model is the natural extension of the functional deformation models studied in the statistical literature for which estimation procedures are provided in Gamboa, Loubes and Maza (2007) while testing issues are tackled in Collier and Dalalyan (2015). Within this framework, statistical inference on deformation models for distributions have been studied first in Czado and Munk (1998), Munk and Czado (1998) and Freitag and Munk (2005), where tests are provided in the case of parametric functions, while the estimation of the parameters is studied in Agulló-Antolín et al. (2015).

In this work, after recalling the model we use in Section 2, we tackle the problem of the estimation of the quantiles of the empirical process of the Wasserstein's variation which is obtained using a bootstrap procedure proved in Section 3. We use these results to provide a goodness of fit test in a general non parametric deformation model. For this, we will use an alignment criterion with respect to the Wasserstein's barycenter of a deformation of the observed distributions. This requires an equivalent of a central limit theorem for the Wasserstein variation of a barycenter of measures in both the general case in Section 4 and under the null assumption (observations are drawn from the deformation model) in Section 5. We obtain the asymptotic distribution of the matching criterion in both cases, with a different normalization under the null assumption (only for the parametric case). Finally proofs are postponed to Section 6.

2. Wasserstein variation and a deformation model for distributions

Assume we observe J samples of n i.i.d random variables $X_{i,j}$ with distribution μ_j, associated to a distribution function $F_j : (c_j, d_j) \mapsto (0,1)$ with density with respect to the Lebesgue measure f_j. Let $\mu_{n,j}$ and $F_{n,j}$ be the empirical measure and empirical distribution function associated to the sample $(X_{i,j})_{1\leqslant i\leqslant n}$.

Many recent work has been conducted to measure the spread or the inner structure of collection of distributions. In this paper we define a notion of variability which relies on the notion of Fréchet mean for the space of probability endowed with the Wasserstein metrics, which we will recall the definition hereafter. First, for $d \geqslant 1$, consider the following set

$$\mathcal{W}_2\left(\mathbb{R}^d\right) = \left\{P \text{ probability on } \mathbb{R}^d \text{ with finite second moment}\right\}.$$

For two probabilities μ and ν in $\mathcal{W}_2\left(\mathbb{R}^d\right)$, we denote by $\Pi(\mu,\nu)$ the set of all probability measures π over the product set $\mathbb{R}^d \times \mathbb{R}^d$ with first (resp. second) marginal μ (resp. ν).

The transportation cost with quadratic cost function, or quadratic transportation cost, between these two measures μ and ν is defined as

$$\mathcal{T}_2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2 \, d\pi(x,y).$$

The quadratic transportation cost allows to endow the set $\mathcal{W}_2\left(\mathbb{R}^d\right)$ with a metric by defining the 2-Wasserstein distance between μ and ν as $W_2(\mu,\nu) = \mathcal{T}_2(\mu,\nu)^{1/2}$. More details on Wasserstein distances and their links with optimal transport problems can be found in Rachev (1984) or Villani (2009) for instance.

When we will consider probabilities in $\mathcal{W}_2(\mathbb{R})$, the Wasserstein distance can be written in a simpler expression as

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

where F (resp. G) is the distribution function associated with μ (resp. ν).

Within this framework, we can define a global measure of separation of a collection of measures μ_j, $j = 1,\ldots,n$ as follows. Given probabilities $\mu_1,\ldots,\mu_J$ on $\mathbb{R}^d$ with finite 2-th moment let

$$V(\mu_1,\ldots,\mu_J) = \inf_{\eta\in\mathcal{W}_2(\mathbb{R}^d)} \left(\frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j,\eta)\right)^{1/2}$$

be the Wasserstein 2-variation of $\mu_1,\ldots,\mu_J$ or the variance of the mu_j's.

In Agueh and Carlier (2011) this minimizer of $\eta \mapsto \frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j,\eta)$ is proved to exist. This measure μ_B is called the barycenter or Fréchet mean of $\mu_1,\ldots,\mu_J$. Hence $V(\mu_1,\ldots,\mu_J) = \left(\frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j,\mu_B)\right)^{1/2}$. The authors prove properties of existence and uniqueness for barycenters of measures in $\mathcal{W}_2\left(\mathbb{R}^d\right)$, while the properties of the empirical version are provided in Boissard, Le Gouic and Loubes (2015).

This quantity which is an extension of the variance for probability distributions is a good candidate to evaluate the concentration of a collection of measures around its Fréchet mean. In particular it can be used to test the existence of a distribution's deformation model, in the sense that all the distributions μ_j would be warped from an unknown distribution template μ by a deformation function $\varphi_j^\star$ which can be parametric or non parametric. More precisely, consider a family of warping functions $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_J$ such that

$$\text{For all } h \in \mathcal{G}_j, h: \begin{array}{c} (c_j, d_j) \to (a, b) \\ x \mapsto h(x) \end{array} \text{ is invertible, increasing,} \tag{$\mathcal{A}\mathbf{1}$}$$
$$\text{and s.t.} -\infty \leqslant a < b \leqslant +\infty, \quad -\infty \leqslant c \leqslant c_j < d_j \leqslant d \leqslant +\infty.$$

We would like to build a goodness-of-fit testing procedure for the following model

$$\text{There exist } (\varphi_1^\star, \dots \varphi_J^\star) \in \mathcal{G} \text{ and } (\varepsilon_{i,j})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant J}} \text{ i.i.d. such that}$$
$$X_{i,j} = \left(\varphi_j^\star\right)^{-1}(\varepsilon_{i,j}) \quad \forall 1 \leqslant j \leqslant J \tag{$\mathcal{H}$}$$

Denote by G the distribution function of ε with law μ with support (a, b), while $G_{n,j}$ is the corresponding empirical version.

We propose to use this Wasserstein 2-variation as a goodness of fit criterion for model ($\mathcal{H}$). Since the true distribution μ is unknown, we first try to invert the warping operator and thus compute for each observation its image through a candidate deformation φ_j,

$$Z_{i,j}(\varphi_j) = \varphi_j(X_{i,j}) \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant J.$$

Note that $Z_{i,j}(\varphi_j) \sim \mu_j(\varphi_j)$ with distribution function (under Assumption $\mathcal{A}\mathbf{1}$) $F_j \circ \varphi_j^{-1} := F_{\varphi_j}$. Now, if we set $\varphi = (\varphi_1, \dots, \varphi_J) \in \mathcal{G}$, then the Fréchet mean of $(\mu_j(\varphi_j))_{1 \leqslant j \leqslant J}$ is the probability $\mu_B(\varphi)$ with quantile function

$$F_B^{-1}(\varphi)(t) = \frac{1}{J} \sum_{k=1}^{J} \varphi_k \circ F_k^{-1}(t),$$

see for instance Agueh and Carlier (2011). We will write $\mu_{n,j}(\varphi_j)$ for the empirical measure on $Z_{i,j}(\varphi_j), 1 \leq i \leq n$ and $\mu_{n,B}(\varphi)$ for the corresponding Fréchet mean. It is important to remark that

$$\text{under } (\mathcal{H}) \quad \mu_B(\varphi^\star) = \mu = \mu_j(\varphi_j^\star), \ \forall 1 \leqslant j \leqslant J.$$

Hence, a natural idea to test whether $\mathcal{H}$ holds, is to consider the Wasserstein 2-variation of the $(\mu_j(\varphi_j)), 1 \leqslant j \leqslant J$, that is to say the minimum alignment of the candidate warped distributions $(\mu_j(\varphi_j))_{1 \leqslant j \leqslant J}$ with respect to their barycenter, namely $\mu_B(\varphi)$. This optimization program corresponds to the minimization in $\varphi \in \mathcal{G}$ of the following theoretical criterion

$$U(\varphi) := V_2^2(\mu_1(\varphi_1), \dots, \mu_J(\varphi_J)) = \frac{1}{J} \sum_{j=1}^{J} W_2^2(\mu_j(\varphi_j), \mu_B(\varphi)).$$

Its empirical version is given by $U_n(\varphi) = \frac{1}{J} W_2^2(\mu_{n,j}(\varphi_j), \mu_{n,B}(\varphi))$. Inference about model $\mathcal{H}$ can be based on the statistic $\inf_{\varphi \in \mathcal{G}} U_n(\varphi)$ which is equal to zero if the distributions can be warped from a common distribution and different otherwise. In the next sections we analyze the behavior of this statistics under different setups.

3. Bootstrap with Wasserstein distance

In this section we present general results on Wasserstein distances that we will apply to estimate the asymptotic distribution of a statistic test based on an alignment with respect to the Wasserstein's barycenter. More precisely, here we consider distributions on $\mathbb{R}^d$ with a moment of order $r \geqslant 1$, that is, distributions in $\mathcal{W}_r(\mathbb{R}^d)$. W_r will denote Wasserstein distance with L_r cost, namely,

$$W_r^r(\mu, \eta) = \inf_{\pi \in \Pi(\mu, \eta)} \int \|y - z\|^r d\pi(y, z),$$

where $\|\cdot\|$ is any norm on $\mathbb{R}^d$. Finally, we write $\mathcal{L}(Z)$ for the law of any random variable Z. We note the abuse of notation in the following, in which W_r is used both for Wasserstein distance on $\mathbb{R}$ and on $\mathbb{R}^d$, but this should not cause much confusion.

The next result shows that the laws of empirical transportation costs are continuous (and even Lipschitz) functions of the underlying distributions.

Theorem 3.1. *Set ν, ν', η probability measures in $\mathcal{W}_r(\mathbb{R}^d)$, $Y_1, \ldots, Y_n$ i.i.d. random vectors with common law ν, $Y_1', \ldots, Y_n'$, i.i.d. with law ν' and write ν_n, ν_n' for the corresponding empirical measures. Then*

$$W_r(\mathcal{L}(W_r(\nu_n, \eta)), \mathcal{L}(W_r(\nu_n', \eta))) \leqslant W_r(\nu, \nu').$$

Our deformation assessment criterion concerns a particular version of the Wasserstein r-variation of distributions $\nu_1, \ldots, \nu_J$ in $\mathcal{W}_r(\mathbb{R}^d)$, that is denoted in its general form by

$$V_r(\nu_1, \ldots, \nu_J) := \inf_{\eta \in \mathcal{W}_r(\mathbb{R}^d)} \left(\frac{1}{J} \sum_{j=1}^{J} W_r^r(\nu_j, \eta) \right)^{1/r}.$$

V_r is just the average distance to the r-barycenter of the set.

It is convenient to note that $V_r^r(\nu_1, \ldots, \nu_J)$ can also be expressed as

$$V_r^r(\nu_1, \ldots, \nu_J) = \inf_{\pi \in \Pi(\nu_1, \ldots, \nu_J)} \int T(y_1, \ldots, y_J) d\pi(y_1, \ldots, y_J), \tag{2}$$

where $\Pi(\nu_1, \ldots, \nu_J)$ denotes the set of probability measures on $\mathbb{R}^d$ with marginals $\nu_1, \ldots, \nu_J$ and $T(y_1, \ldots, y_J) = \min_{z \in \mathbb{R}^d} \frac{1}{J} \sum_{j=1}^{J} \|y_j - z\|^r$.

Here we are interested in empirical Wasserstein r-variations, namely, the r-variations computed from the empirical measures $\nu_{n_j,j}$ coming from independent samples $Y_{1,j},\dots,Y_{n_j,j}$ of i.i.d. random variables with distribution ν_j. Note that in this case problem (2) is a linear optimisation problem for which a minimizer always exists.

As before, we consider the continuity of the law of empirical Wasserstein r-variations with respect to the underlying probabilities. This is covered in the next result.

Theorem 3.2. *With the above notation*

$$W_r^r(\mathcal{L}(V_r(\nu_{n_1,1},\dots,\nu_{n_J,J})),\mathcal{L}(V_r(\nu'_{n_1,1},\dots,\nu'_{n_J,J}))) \leqslant \frac{1}{J}\sum_{j=1}^{J} W_r^r(\nu_j,\nu'_j).$$

A useful consequence of the above results is that empirical Wasserstein distances or r-variations can be bootstrapped under rather general conditions. To be more precise, we take in Theorem 3.1 $\nu'=\nu_n$, the empirical measure on $Y_1,\dots,Y_n$ and consider a bootstrap sample $Y_1^*,\dots,Y_{m_n}^*$ of i.i.d. (conditionally given $Y_1,\dots,Y_n$) observations with common law ν_n. We write $\nu_{m_n}^*$ for the empirical measure on $Y_1^*,\dots,Y_{m_n}^*$ and $\mathcal{L}^*(Z)$ for the conditional law of Z given $Y_1,\dots,Y_n$. Proposition 3.1 now reads

$$W_r(\mathcal{L}^*(W_r(\nu_{m_n}^*,\nu)),\mathcal{L}(W_r(\nu_{m_n},\nu))) \leqslant W_r(\nu_n,\nu).$$

Hence, if $W_r(\nu_n,\nu)=O_{\mathbb{P}}(1/r_n)$ for some sequence $r_n>0$ such that $r_{m_n}/r_n\to 0$ as $n\to\infty$, then, using that $W_r(\mathcal{L}(aX),\mathcal{L}(aY))=aW_r(\mathcal{L}(X),\mathcal{L}(Y))$ for $a>0$, we see that

$$W_r(\mathcal{L}^*(r_{m_n}W_r(\nu_{m_n}^*,\nu)),\mathcal{L}(r_{m_n}W_r(\nu_{m_n},\nu))) \leqslant \frac{r_{m_n}}{r_n} r_n W_r(\nu_n,\nu)\to 0 \tag{3}$$

in probability.

If in addition $r_nW_r(\nu_n,\nu)\rightharpoonup\gamma(\nu)$ for a distribution $\gamma(\nu)$ then

$$r_{m_n}W_r(\nu_{m_n},\nu)\rightharpoonup\gamma(\nu) \tag{4}$$

which entails that if $\hat{c}_n(\alpha)$ denotes the α quantile of the conditional distribution $\mathcal{L}^*(r_{m_n}W_r(\nu_{m_n}^*,\nu))$ then under some regularity conditions on the distribution $\gamma(\nu)$

$$\mathbb{P}\left(r_nW_r(\nu_n,\nu)\leqslant\hat{c}_n(\alpha)\right)\to\alpha \quad \text{as } n\to\infty. \tag{5}$$

We conclude in this case that the quantiles of $r_nW_r(\nu_n,\nu)$ can be consistently estimated by the bootstrap quantiles, that is, the conditional quantiles of the quantity $r_{m_n}W_r(\nu_{m_n}^*,\nu)$ (which, in turn, can be approximated through Monte-Carlo simulation).
As an example, if $d=1$ and $r=2$, under integrability and smoothness assumptions on ν we have $\sqrt{n}W_2(\nu_n,\nu)\rightharpoonup\left(\int_0^1\frac{B^2(t)}{f^2(F^{-1}(t))}dt\right)^{1/2}$, where f and F^{-1} are the density and the quantile function of ν.

In the same way the Wasserstein r-variation can also be bootstrapped as soon as a limit theorem exists as in (4). This is the purpose of the following sections in the framework of a deformation model.

4. Non parametric model for deformations

Recall that we consider μ_j's that may be warped from a common unknown template by deformations $\varphi_j^\star$'s. We consider the Wasserstein 2-variation for a given set of candidate deformation function φ, namely $U(\varphi) := V_2^2(\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J))$ and its empirical version $U_n(\varphi)$. We prove a Central Limit Theorem for this quantity in order to design a goodness of fit test for the deformation model.

4.1. Central Limit Theorem for Wasserstein's barycenter variations

We provide in the section a CLT for $\inf_{\varphi\in\mathcal{G}} U_n(\varphi)$ under the following set of assumptions.

$$\text{For all } j,\ F_j \text{ is } C^2 \text{ on } (c_j;d_j),\ f_j(x)>0 \text{ if } x\in(c_j;d_j) \text{ and} \qquad (\mathcal{A}\mathbf{2})$$
$$\sup_{c_j<x<d_j} \frac{F_j(x)(1-F_j(x))f_j'(x)}{f_j(x)^2} < \infty.$$

$$\text{For some } q>1 \text{ and all } 1\leqslant j\leqslant J, \quad \int_0^1 \frac{(t(1-t))^{\frac{q}{2}}}{\left(f_j\left(F_j^{-1}(t)\right)\right)^q} dt < +\infty \qquad (\mathcal{A}\mathbf{3})$$

For q as in $\mathcal{A}\mathbf{3}$, we set $p_0 = \max\left(\frac{q}{q-1}, 2\right)$ and define on $\mathcal{H}_j = C^1(c_j;d_j)\cap L^{p_0}(X_j)$ the norm $\|h_j\|_{\mathcal{H}_j} = \sup_{(c_j;d_j)} |h_j'| + \mathbb{E}\left[|h_j(X_j)|^{p_0}\right]^{\frac{1}{p_0}}$, and on the product space $\mathcal{H}_1\times\cdots\times\mathcal{H}_J$, $\|h\|_{\mathcal{H}} = \sum_{j=1}^J \|h_j\|_{\mathcal{H}_j}$. The we make the following additional assumptions.

$$\mathcal{G}_j \subset \mathcal{H}_j \text{ is compact for } \|\cdot\|_{\mathcal{H}_j} \text{ and } \sup_{h\in\mathcal{G}_j} \left|h'(x_n^h) - h'(x)\right| \underset{\sup_{h\in\mathcal{G}_j} |x_n^h - x|\to 0}{\to} 0. \qquad (\mathcal{A}\mathbf{4})$$

$$\text{for some } r>4 \text{ and } 1\leqslant j\leqslant J, \quad \mathbb{E}\left[|X_j|^r\right] < \infty \qquad (\mathcal{A}\mathbf{5})$$

$$\text{for some } r>\max(4,p_0) \text{ and } 1\leqslant j\leqslant J, \quad \mathbb{E}\left[\sup_{h\in\mathcal{G}_j} |h(X_j)|^r\right] < \infty \qquad (\mathcal{A}\mathbf{6})$$

Under $\mathcal{A}\mathbf{1}$ to $\mathcal{A}\mathbf{6}$, we are able to provide the asymptotic distribution of $\inf_{\varphi\in\mathcal{G}} \sqrt{n}U_n(\varphi)$. It is convenient at this point to give some explanation about the meaning of these assumptions. $\mathcal{A}\mathbf{2}$ is a is a regularity condition on the distributions of the $X_j's$ (it holds, for instance, for Gaussian or Pareto distributions) required for strong approximation of the quantile process, see Csörgő (1983) for details. The integrability condition $\mathcal{A}\mathbf{3}$ is satisfied by the Gaussian distribution if $q<2$, see, e.g., Rajput (1972). $\mathcal{A}\mathbf{4}$ is related to the regularity of the deformation functions. Finally, $\mathcal{A}\mathbf{5}$ and $\mathcal{A}\mathbf{6}$ are moment assumptions on the (possibly warped) observations.

Theorem 4.1. *Under Assumptions* $\mathcal{A}$**1** *to* $\mathcal{A}$**6**

$$\sqrt{n}\Big(\inf_{\varphi\in\mathcal{G}} U_n(\varphi) - \inf_{\varphi\in\mathcal{G}} U(\varphi)\Big)$$
$$\rightharpoonup \inf_{\varphi\in\Gamma} \frac{2}{J}\sum_{j=1}^{J}\int_0^1 \varphi_j'\circ F_j^{-1}\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1} - F_B^{-1}(\varphi)),$$

where $\Gamma = \{\varphi\in\mathcal{G} : U(\varphi) = \inf_{\phi\in\mathcal{G}} U(\phi)\}$ *and* $(B_j)_{1\leqslant j\leqslant J}$ *are independent Brownian bridges.*

A proof of Theorem 4.1 is given in the Appendix below. We note that continuity of U is follows easily from the choice of the norm on $\mathcal{G}$. Recall that $\mathcal{G}$ is compact and, consecuently, $\inf_{\varphi\in\mathcal{G}} U(\varphi)$ is attained. Hence, Γ is a nonempty closed subset of $\mathcal{G}$ (in particular, it is also a compact set). We note further that the random variables $\int_0^1 \varphi_j'\circ F_j^{-1}\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1} - F_B^{-1}(\varphi))$ are centered Gaussian, with covariance

$$\int_{[0,1]^2}(\min(s,t)-st)\frac{\varphi_j'(F_j^{-1}(t))}{f_j\big(F_j^{-1}(t)\big)}(\varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t))$$
$$\times\quad \frac{\varphi_j'(F_j^{-1}(s))}{f_j\big(F_j^{-1}(s)\big)}(\varphi_j(F_j^{-1}(s)) - F_B^{-1}(\varphi)(s))dsdt.$$

In particular, if U has a unique minimizer the limiting distribution in Theorem 4.1 is normal. However, our result works in more generality, even without uniqueness assumptions.

We remark also that although we have focused for simplicity on the case of samples of equal size, the case of different sample sizes, n_j, $j = 1,\dots,J$, can also be handled with straightforward changes. If we assume

$$\forall j : n_j \to +\infty \text{ and } \frac{n_j}{n_1+\cdots+n_J} \to (\gamma_j)^2 > 0, \tag{6}$$

then the result can be restated as

$$\sqrt{\frac{n_1\dots n_J}{(n_1+\cdots+n_J)^{J-1}}}\Big(\inf_{\mathcal{G}} U_{n_1,\dots,n_J} - \inf_{\mathcal{G}} U\Big) \rightharpoonup \inf_{\Gamma}\frac{2}{J}\sum_{j=1}^{J}\widetilde{S}_j,$$

where $U_{n_1,\dots,n_J}$ denotes the empirical Wasserstein variation computed from the samples and $\widetilde{S}_j(\varphi) = \big(\Pi_{p\neq j}\gamma_p\big)\int_0^1 \varphi_j'\circ F_j^{-1}\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1} - F_B^{-1}(\varphi))$.

As a final remark in this section we note that in the case where $\mathcal{H}$ holds, we have $\varphi_j\circ F_j^{-1} = F_B^{-1}(\varphi)$ for each $\varphi\in\Gamma$. Thus, the result of Theorem 4.1 becomes

$$\inf_{\varphi\in\mathcal{G}}\sqrt{n}U_n(\varphi) \rightharpoonup 0.$$

Hence, in this case we have to refine our study to understand well the behavior of $\inf_{\mathcal{G}} U_n$ when n tends to infinity. This is what we consider in the next section. In this case we restrict ourselves to a to a semiparametric warping model where μ is unknown but where the deformations are indexed by a parametric family.

4.2. *A testing procedure*

In the non parametric deformation model, statistical inference is based on the minimal Wasserstein variation

$$v_n^2 := \inf_{\varphi \in \mathcal{G}} U_n(\varphi),$$

where $\mu_{n,j}(\varphi)$ denotes the empirical measure on $Z_{1,j}(\varphi), \dots, Z_{n,j}(\varphi)$, where $Z_{i,j}(\varphi) = \varphi_j^{-1}(X_{i,j})$ and $X_{1,j}, \dots, X_{n,j}$ are independent i.i.d. samples from μ_j. Consider v_n', the corresponding version obtained from samples with underlying distributions μ_j', and denote by $\mathcal{L}(v_n)$ (reps. $\mathcal{L}(v_n')$) the law of the random variable v_n (resp. v_n').

Then, the following result holds, setting $\|\varphi_j\|_\infty = sup_{x \in (c_j;d_j)} |\varphi_j(x)|$.

Theorem 4.2. *Under Assumption $\mathcal{A}$**1**, if for all j $\mathcal{G}_j \subset \mathcal{C}^1(c_j; d_j)$ and $\sup_{\varphi \in \mathcal{G}} \|\varphi_j'\|_\infty < \infty$, then*

$$W_2^2(\mathcal{L}(v_n), \mathcal{L}(v_n')) \leqslant \sup_{\varphi \in \mathcal{G}} \|\varphi_j'\|_\infty^2 \frac{1}{J} \sum_{j=1}^{J} W_2^2(\mu_j, \mu_j').$$

Now consider bootstrap samples $X_{1,j}^*, \dots, X_{m_n,j}^*$ of i.i.d. observations sampled from $\mu_{n,j}$, write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \dots, X_{m_n,j}^*$ (conditionally to the $X_{1,j}, \dots, X_{n,j}$) and denote $V_2^2(\mu_{m_n,1}^*(\varphi), \dots, \mu_{m_n,J}^*(\varphi)) = U_{m_n}^*(\varphi)$. Then we get

Corollary 4.3. *If $m_n \to \infty$, and $m_n/\sqrt{n} \to 0$, then under Assumptions $\mathcal{A}$**1** to $\mathcal{A}$**6**, and if $\inf_{\mathcal{G}} U > 0$, writing γ for the limit distribution in Theorem 4.1, we have that*

$$\mathcal{L}^* \left(\sqrt{m_n} \left(\inf_{\mathcal{G}} U_{m_n}^* - \inf_{\mathcal{G}} U \right) \right) \rightharpoonup \gamma$$

in probability. In particular, if $\hat{c}_n(\alpha)$ denotes the conditional (given the $X_{i,j}$'s) α-quantile of $\sqrt{m_n}\left(\inf_{\mathcal{G}} U_{m_n}^* - \inf_{\mathcal{G}} U\right)$ then

$$\mathbb{P}\left(\sqrt{n} \left(\inf_{\mathcal{G}} U^n - \inf_{\mathcal{G}} U \right) \leqslant \hat{c}_n(\alpha) \right) \to \alpha. \tag{7}$$

Since under $\mathcal{H}$ we get $\inf_{\varphi \in \mathcal{G}} \sqrt{n} U_n(\varphi) \rightharpoonup 0$, we can not obtain a goodness of fit test in the non parametric model. Yet a test can be designed switching the null hypothesis. Hence set for $\Delta_0 > 0$ set

$$\inf_{\varphi \in \mathcal{G}} U(\varphi) = \Delta_0 \tag{$\mathcal{H}_0^1$}$$

$$\inf_{\varphi \in \mathcal{G}} U(\varphi) < \Delta_0 \tag{$\mathcal{H}_1^1$}$$

The test statistic in this case is $\mathcal{U}_n(\Delta_0) := \sqrt{n}(\inf_{\mathcal{G}} U_n - \Delta_0)$. Then, under assumptions of Corollary 4.3 ($\mathcal{A}$**1** to $\mathcal{A}$**6**), if $\hat{c}_n(\alpha)$ denotes the conditional (given the $X_{i,j}$'s) α-quantile of the bootstrap version $\sqrt{m_n}\left(\inf_{\mathcal{G}} U_{m_n}^* - \Delta_0\right)$, under $\mathcal{H}_0^1$

$$\mathbb{P}(\mathcal{U}_n(\Delta_0) \leqslant \hat{c}_n(\alpha)) \to \alpha,$$

which gives the asymptotic level of the reject region $\{\mathcal{U}_n(\Delta_0) \leqslant \hat{c}_n(\alpha)\}$, where $\hat{c}_n(\alpha)$ can be computed using a Monte-Carlo method.

5. A parametric model for deformations

5.1. Estimation of the deformation model

In many cases, deformation functions can be made more specific in the sense that they follow a known shape depending on parameters that are different for each sample. So consider the parametric model $\theta^\star = (\theta_1^\star, \ldots, \theta_J^\star)$ such that $\varphi_j^\star = \varphi_{\theta_j^\star}$, for all $j = 1, \ldots, J$. Each $\theta_j^\star$ represents the warping effect that undergoes the j^{th} sample, which must be removed to recover the unknown distribution by inverting the warping operator. So Assumption $\mathcal{H}$ becomes

$$X_{i,j} = \varphi_{\theta_j^\star}^{-1}(\varepsilon_{i,j}), \text{ for all } 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant J.$$

Hence, from now on, we will consider the following family of deformations, indexed by a parameter $\lambda \in \Lambda \subset \mathbb{R}^p$:

$$\begin{array}{rcl} \varphi : \Lambda \times (c;d) & \to & (a,b) \\ (\lambda, x) & \mapsto & \varphi_\lambda(x) \end{array}$$

Thus, the functions U and U_n are now defined on $\Theta = \Lambda^J$, and the criterion of interest becomes $\inf_{\lambda \in \Theta} U(\lambda)$. We also use the simplified notation $\mu_j(\theta_j)$ instead of $\mu_j(\varphi_{\theta_j})$, $F_B(\theta)$ for $F_B(\varphi_{\theta_1}, \ldots, \varphi_{\theta_J})$ and similarly for the empirical versions. Throughout this section *we assume that model $\mathcal{H}$ holds*. This means, in particular, that the d.f.'s of the samples, F_j, satisfy $F_j = G \circ \varphi_{\theta_j^*}$, with G the d.f. of the $\varepsilon_{i,j}$'s.

For the analysis of this setup, we adapt Assumptions $\mathcal{A}$**1** to $\mathcal{A}$**6**, replacing them by the following versions.

$$\text{For all } \lambda \in \Lambda, \varphi_\lambda : \begin{array}{c} (c;d) \to (a;b) \\ x \mapsto \varphi_\lambda(x) \end{array} \text{ is invertible, increasing,} \tag{A1}$$
$$\text{and s.t.} -\infty \leqslant a < b \leqslant +\infty, \quad -\infty \leqslant c \leqslant c_j < d_j \leqslant d \leqslant +\infty.$$

We replace $\mathcal{A}$**2** by: G is C^2 with $G'(x) = g(x) > 0$ on (a,b) and

$$\sup_{a<x<b} \frac{G(x)(1-G(x))\,g'(x)}{g(x)^2} < \infty \tag{A2}$$

Now, instead of $\mathcal{A}$**3** to $\mathcal{A}$**5** we assume

$$\varphi \text{ is continuous w.r.t. } x \text{ and } \lambda \tag{A3}$$
$$\forall \lambda \in \Lambda,\ \varphi_\lambda \text{ is } C^1 \text{ with respect to } x,\ \Lambda \text{ is compact}$$

$$d\varphi \text{ is bounded on } \Lambda \times [c_j; d_j] \text{ and continuous with respect to } \lambda$$
$$\text{and } \sup_{\lambda \in \Lambda} \left| d\varphi_\lambda(x_n^\lambda) - d\varphi_\lambda(x) \right| \xrightarrow{\sup_{\lambda \in \Lambda} |x_n^\lambda - x| \to 0} 0. \tag{A4}$$

$$\forall 1 \leqslant j \leqslant J \quad \mathbb{E}\left[|X_j|^r\right] < \infty \text{ for some } r > 4 \tag{A5}$$

Here d is the derivation operator w.r.t. x, while ∂ will be the derivation operator w.r.t. λ. Finally $\mathcal{A}\mathbf{6}$ becomes

$$\forall 1 \leqslant j \leqslant J \quad \mathbb{E}\left[\sup_{\lambda\in\Lambda} |\varphi_\lambda(X_j)|^r\right] < \infty \text{ for some } r > 4 \tag{A6}$$

Note that Assumption **A6** implies that ε has a moment of order $r > 4$ and also that Assumption $\mathcal{A}\mathbf{3}$ becomes simpler in a parametric model which does not require a particular topology.

We impose as identifiability condition,

$$U \text{ has a unique minimizer, } \theta^\star, \text{ that belongs to the interior of } \Lambda. \tag{A7}$$

Note that, equivalently, this means that θ^* is the unique zero of U, since we are assuming that $\mathcal{H}$ holds.

Now, to get sharper result about the convergence of $\inf_{\theta\in\Theta} U^n(\theta)$, one has to add the following assumptions, first on the deformation functions.

$$\forall 1 \leqslant j \leqslant J \quad \varphi^{-1}_{\theta^\star_j} \text{ is } C^1 \text{ w.r.t. } x \text{ and } d\varphi^{-1}_{\theta^\star_j} \text{ is bounded on } [a,b] \tag{A8}$$

$$\varphi \text{ is } C^2 \text{ w.r.t. } x \text{ and } \lambda$$

$$\forall 1 \leqslant j \leqslant J \quad \mathbb{E}\left[\sup_{\lambda\in\Lambda}\left|\partial^2\varphi_\lambda\left(\varphi^{-1}_{\theta^\star_j}(\varepsilon)\right)\right|^2\right] < \infty \tag{A9}$$

As said for Assumption $\mathcal{A}\mathbf{3}$, the following one is more restrictive on the tail of the distribution of ε, excluding the Gaussian case. Examples of such variables with unbounded support are given in del Barrio, Deheuvels and van de Geer (2007) p.76. Note that distributions with compact support and strictly positive, continuous density satisfy this assumption.

$$\int_0^1 \frac{t(1-t)}{g^2\left(G^{-1}(t)\right)} dt < \infty \tag{A10}$$

Set $U_n(\theta) = U_n(\varphi_\theta) = V_2^2(\mu_{n,1}(\varphi_{\theta_1}), \ldots, \mu_{n,J}(\varphi_{\theta_J}))$ and consider the estimator

$$\widehat{\theta}^n \in \arg\min_{\theta\in\Theta} U^n(\theta).$$

The results in this section are stated in the case where Λ is a subset of $\mathbb{R}$. However they are still true if $\Lambda \subset \mathbb{R}^p$ with corresponding changes. The following result implies that $\widehat{\theta}^n$ is a good candidate to estimate $\theta^\star$. It is a simple consequence of the continuity of U and the uniform convergence in probability of U_n to U, as shown in the proof of Theorem 4.1.

Proposition 5.1. *Under* **A1** *to* **A7**, *then*

$$\widehat{\theta}^n \to \theta^\star \text{ in probability.}$$

We can refine this result by making the following additional assumption,

$$R_j := \partial\varphi_{\theta_j^*} \circ \varphi_{\theta_j^*}^{-1} \text{ is continuous and bounded on } [a,b],\ 1 \le j \le J. \qquad \textbf{(TCL)}$$

Define now $\Phi = [\Phi_{i,j}]_{1\le i,j\le J}$ with

$$\Phi_{i,j} = -\frac{2}{J^2}\langle R_i, R_j\rangle_\mu,\ i \ne j; \quad \Phi_{i,i} = \frac{2(J-1)}{J^2}\|R_i\|_\mu, \tag{8}$$

where $\|\cdot\|_\mu$ and $\langle\cdot,\cdot\rangle_\mu$ denote norm and inner product, respectively, in $L^2(\mu)$. Φ is a symmetric, positive semidefinite matrix. To see this, consider $x \in \mathbb{R}^J$ and note that

$$\begin{aligned} x'\Phi x &= \frac{2}{J^2}\int \big(\sum_i (J-1)x_i^2R_i^2 - 2\sum_{i<j} x_ix_jR_iR_j\big)d\mu \\ &= \frac{2}{J^2}\int \sum_{i<j}(x_iR_i - x_jR_j)^2 d\mu \ge 0. \end{aligned}$$

In fact, Φ is positive definite, hence invertible, unless all the R_i are proportional μ-a.s.. Now, we can state the following Central limit Theorem.

Proposition 5.2. *Under Assumptions* **A1** *to* **A9** *and* **TCL**, *if, in addition,* Φ *is invertible, then*

$$\sqrt{n}(\hat\theta^n - \theta^\star) \rightharpoonup \Phi^{-1}Y,$$

where $Y \overset{d}{=} (Y_1,\dots,Y_J)$ *with*

$$Y_j = \frac{2}{J}\int_0^1 R_j \circ G^{-1}\frac{\tilde B_j}{g\circ G^{-1}},$$

$\tilde B_j = B_j - \frac{1}{J}\sum_{k=1}^J B_k$ *and* $(B_j)_{1\leqslant j\leqslant J}$ *independent Brownian bridges.*

We note that, while, for simplicity, we have formulated Proposition 5.1 assuming that the deformation model holds, a similar version can be proved (with some additional assumptions and changes in Φ) in the case when the model is false and θ^* is not the *true* parameter, but the one that gives the best (but imperfect) alignment.

Remark 1. *The indentifiability condition* **A7** *can be too strong to be realistic. Actually, for some deformation models it could happen that* $\varphi_\theta \circ \varphi_\eta = \varphi_{\theta*\eta}$ *for some* $\theta * \eta \in \Theta$. *In this case, if* $X_{i,j} = \varphi_{\theta_j^*}^{-1}(\varepsilon_{i,j})$ *with* $\varepsilon_{i,j}$ *i.i.d., then, for any* θ, $X_{i,j} = \varphi_{\theta*\theta_j^*}^{-1}(\tilde\varepsilon_{i,j})$ *with* $\tilde\varepsilon_{i,j} = \varphi_\theta(\varepsilon_{i,j})$ *which are also i.i.d. and, consequently,* $(\theta*\theta_1^*,\dots,\theta*\theta_J^*)$ *is also a zero of* U. *This applies, for instance, to location and scale models. A simple fix to this issue is to select one of the signals as the reference, say the* J*-th signal, and assume that* θ_J^* *is known (since it can be, in fact, chosen arbitrarily). The criterion function becomes then* $\tilde U(\theta_1,\dots,\theta_{J-1}) = U(\theta_1,\dots,\theta_{J-1},\theta_J^*)$. *One could then make the (more realistic) assumption that* $\tilde\theta^* = (\theta_1^*,\dots,\theta_{J-1}^*)$ *is the unique zero of* $\tilde U$ *and base the* *analysis on* $\tilde{U}_n(\theta_1,\dots,\theta_{J-1}) = U_n(\theta_1,\dots,\theta_{J-1},\theta_J^*)$ *and* $\hat{\tilde{\theta}}^n = \arg\min_{\tilde{\theta}} \tilde{U}_n(\tilde{\theta})$. *The results in this section can be adapted almost verbatim to this setup. Proposition 5.2 holds, namely,* $\sqrt{n}(\hat{\tilde{\theta}}^n - \tilde{\theta}^*) \rightharpoonup \tilde{\Phi}^{-1}\tilde{Y}$, *with* $\tilde{Y} \stackrel{d}{=} (Y_1,\dots,Y_{J-1})$ *and* $\tilde{\Phi} = [\Phi_{i,j}]_{1\le i,j\le J-1}$. *We note further that invertibility of* $\tilde{\Phi}$ *is almost granted. In fact, arguing as above, we see that*

$$x'\tilde{\Phi}x = \frac{2}{J^2}\int\Big(\sum_{1\le i<j\le J-1}(x_iR_i - x_jR_j)^2 + \sum_{1\le i\le J-1} x_i^2R_i^2\Big)d\mu \ge 0$$

and $\tilde{\Phi}$ *is positive definite unless* $R_i = 0$ μ*-c.s. for* $i = 1,\dots,J-1$.

5.2. Asymptotic behavior of Wasserstein's variation under the null assumption

Here we are able to specify the rate of convergence of $\inf_{\theta\in\Theta} U_n(\theta)$ to zero when $\mathcal{H}$ holds, and to provide the asymptotic distribution of this statistic.

Theorem 5.3. *Under assumptions* **A1** *to* **A10**, **TCL** *and invertibility of* Φ,

$$n\inf_{\theta\in\Theta} U_n(\theta) \rightharpoonup \frac{1}{J}\sum_{j=1}^J \int_0^1 \Big(\frac{\tilde{B}_j}{g\circ G^{-1}}\Big)^2 - \frac{1}{2}Y'\Phi^{-1}Y$$

with $Y = (Y_1,\dots,Y_J)$, $Y_j = \frac{2}{J}\int_0^1 R_j\circ G^{-1}\frac{\tilde{B}_j}{g\circ G^{-1}}$, $\tilde{B}_j = B_j - \frac{1}{J}\sum_{k=1}^J B_k$ *and* $(B_j)_{1\leqslant j\leqslant J}$ *independent Brownian bridges.*

Turning back to our goal of assessment of the deformation model $\mathcal{H}$ based on the observed value of $\inf_{\theta\in\Theta} U^n(\theta)$, Theorem 5.3 gives some insight into the threshold levels for rejection of $\mathcal{H}$. However, the limiting distribution still depends on unknown objects and designing a tractable test requires to estimate the quantiles of this distribution. This will be achieved in the next subsection.

5.3. Goodness of fit test for parametric deformation model

Now consider the parametric deformation model and note that the inference about it is based on the minimal Wasserstein variation indexed by $\theta\in\Theta$

$$v_n^2 := \inf_{\theta\in\Theta} V_2^2(\mu_{n,1}(\theta),\dots,\mu_{n,J}(\theta)) = \inf_{\Theta} U_n,$$

where $\mu_{n,j}(\theta)$ denotes the empirical measure on $Z_{1,j}(\theta),\dots,Z_{n,j}(\theta)$, $Z_{i,j}(\theta) = \varphi_{\theta_j}^{-1}(X_{i,j})$ and $X_{1,j},\dots,X_{n,j}$ are independent i.i.d. samples from μ_j. We consider v_n', the corresponding version obtained from samples with underlying distributions μ_j', and denote by $\mathcal{L}(v_n)$ (resp. $\mathcal{L}(v_n')$) the law of the random variable v_n (resp. v_n').

Then, mimicking the proof of Theorem 4.2 ,we are able to prove the following result.

Theorem 5.4. *Under Assumptions* **A1**, **A3** *and* **A4**

$$W_2^2(\mathcal{L}(v_n), \mathcal{L}(v_n')) \leqslant \sup_{x\in(c;d),\lambda\in\Lambda} |d\varphi_\lambda(x)|^2 \frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j, \mu_j').$$

Now consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of i.i.d. observations sampled from μ_j^n, write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \ldots, X_{m_n,j}^*$ (conditionally to the $X_{1,j}, \ldots, X_{n,j}$) and denote $V_2^2(\mu_{m_n,1}^*(\theta), \ldots, \mu_{m_n,J}^*(\theta)) = U_{m_n}^\star(\theta)$.

Corollary 5.5. *If $m_n \to \infty$, and $m_n/n \to 0$, then under Assumptions* **A1** *to* **A10**, **TCL** *and writing $\gamma(G;\theta^*)$ for the limit distribution in Theorem 5.3, we have that*

$$\mathcal{L}^*\left(m_n \inf_\Theta U_{m_n}^*\right) \rightharpoonup \gamma(G;\theta^*)$$

in probability. In particular, if $\hat{c}_n(\alpha)$ denotes the conditional (given the $X_{i,j}$'s) α-quantile of $m_n \inf_\Theta U_{m_n}^*$ then if the quantile function of $\gamma(G;\theta^*)$ is continuous w.r.t α

$$\mathbb{P}\left(n \inf_\Theta U^n \leqslant \hat{c}_n(\alpha)\right) \to \alpha. \tag{9}$$

In the semi parametric model, we can now provide a goodness of fit procedure. Under Assumptions of Theorem 5.3 (**A1** to **A10** and **TCL**) one can test the null assumption

$$\inf_{\theta\in\Theta} U(\theta) = 0 \tag{$\mathcal{H}_0$}$$

versus its complementary denoted by $\mathcal{H}_1$.

In this case the test statistic is $n\inf_\Theta U_n$ and one can get the asymptotic level of a reject region of the form $\{n\inf_\Theta U_n > \lambda_n\}$ by using Corollary 5.5.

More precisely, consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of i.i.d. observations sampled from $\mu_{n,j}$, and write $U_{m_n}^*(\theta)$ for the corresponding criterion. Then, if $\hat{c}_n(\alpha)$ denotes the conditional (given the $X_{i,j}$'s) $(1-\alpha)$-quantile of $m_n \inf_\Theta U_{m_n}^*$

$$\mathbb{P}\left(n \inf_{\theta\in\Theta} U_n(\theta) > \hat{c}_n(\alpha)\right) \to \alpha.$$

Thus $\{n\inf_{\theta\in\Theta} U_n(\theta) > \hat{c}_n(\alpha)\}$ will be a reject region of asymptotic level α, and $\hat{c}_n(\alpha)$ can be computed using a Monte-Carlo method.

6. Appendix

6.1. Proofs of goodness of fit and bootstrap results

Proof of Theorem 3.1. We set $T_n = W_r(\nu_n, \eta)$ and $T_n' = W_r(\nu_n', \eta)$ and $\Pi_n(\eta)$ for the set of probabilities on $\{1,\ldots,n\}\times\mathbb{R}^d$ with first marginal equal to the discrete uniform distribution on $\{1,\ldots,n\}$ and second marginal equal to η and note that we have $T_n = \inf_{\pi\in\Pi_n(\eta)} a(\pi)$ if we denote

$$a(\pi) = \left(\int_{\{1,\ldots,n\}\times\mathbb{R}^d} \|Y_i - z\|^r d\pi(i,z)\right)^{1/r}.$$

We define similarly $a'(\pi)$ from the Y'_i sample to get $T'_n = \inf_{\pi\in\Pi_n(\eta)} a'(\pi)$. But then, using the inequality $|\|a\| - \|b\|| \leqslant \|a-b\|$,

$$|a(\pi) - a'(\pi)| \leqslant \left(\int_{\{1,\dots,n\}\times\mathbb{R}^d} \|Y_i - Y'_i\|^r d\pi(i,z)\right)^{1/r} = \left(\frac{1}{n}\sum_{i=1}^n \|Y_i - Y'_i\|^r\right)^{1/r}$$

This implies that

$$|T_n - T'_n|^r \leqslant \frac{1}{n}\sum_{i=1}^n \|Y_i - Y'_i\|^r.$$

If we take now (Y, Y') to be an optimal coupling of ν and ν', so that $E\left[\|Y-Y'\|^r\right] = W_r^r(\nu,\nu')$ and $(Y_1, Y'_1), \dots, (Y_n, Y'_n)$ to be i.i.d. copies of (Y, Y') we see that for the corresponding realizations of T_n and T'_n we have

$$\mathbb{E}\left[|T_n - T'_n|^r\right] \leqslant \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|Y_i - Y'_i\|^r\right] = W_r(\nu,\nu')^r.$$

But this shows that $W_r(\mathcal{L}(T_n), \mathcal{L}(T'_n)) \leqslant W_r(\nu,\nu')$, as claimed.

□

Proof of Theorem 3.2. We write $V_{r,n} = V_r(\nu_{n_1,1},\dots,\nu_{n_J,J})$ and $V'_{r,n} = V_r(\nu'_{n_1,1},\dots,\nu'_{n_J,J})$. We note that

$$V_{r,n}^r = \inf_{\pi\in\Pi(U_1,\dots,U_J)} \int T(i_1,\dots,i_J) d\pi(i_1,\dots,i_J),$$

where U_j is the discrete uniform distribution on $\{1,\dots,n_j\}$ and $T(i_1,\dots,i_J) = \min_{z\in\mathbb{R}^d} \frac{1}{J}\sum_{j=1}^J \|Y_{i_j,j} - z\|^r$. We write $T'(i_1,\dots,i_J)$ for the equivalent function computed from the $Y'_{i,j}$'s. Hence we have

$$|T(i_1,\dots,i_J)^{1/r} - T'(i_1,\dots,i_J)^{1/r}|^r \leqslant \frac{1}{J}\sum_{j=1}^J \|Y_{i_j,j} - Y'_{i_j,j}\|^r,$$

which implies

$$\begin{aligned}
&\left|\left(\int T(i_1,\dots,i_J) d\pi(i_1,\dots,i_J)\right)^{1/r} - \left(\int T(i_1,\dots,i_J) d\pi(i_1,\dots,i_J)\right)^{1/r}\right|^r \\
&\leqslant \int \frac{1}{J}\sum_{j=1}^J \|Y_{i_j,j} - Y'_{i_j,j}\|^r d\pi(i_1,\dots,i_J) \\
&= \frac{1}{J}\sum_{j=1}^J \int \|Y_{i_j,j} - Y'_{i_j,j}\|^r d\pi(i_1,\dots,i_J) = \frac{1}{J}\sum_{j=1}^J \left(\frac{1}{n_j}\sum_{i=1}^{n_j} \|Y_{i,j} - Y'_{i,j}\|^r\right)
\end{aligned}$$

So,

$$|V_{r,n} - V'_{r,n}|^r \leqslant \frac{1}{J}\sum_{j=1}^J \left(\frac{1}{n_j}\sum_{i=1}^{n_j} \|Y_{i,j} - Y'_{i,j}\|^r\right).$$

If we take (Y_j, Y_j') to be an optimal coupling of ν_j and ν_j' and $(Y_{1,j}, Y_{1,j}'), \ldots, (Y_{n_j,j}, Y_{n_j,j}')$ to be i.i.d. copies of (Y_j, Y_j'), for $j = 1, \ldots, J$, then we obtain

$$\mathbb{E}\left[|V_{r,n} - V_{r,n}'|^r\right] \leqslant \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\mathbb{E}\left[\|Y_{i,j} - Y_{i,j}'\|^r\right]\right) = \frac{1}{J}\sum_{j=1}^{J} W_r^r(\nu_j, \nu_j').$$

The conclusion follows.

$\square$

Proof of Theorem 4.2. We can mimic the argument in the proof of Theorem 3.2 to get an upper bound on the Wasserstein distance between the laws of v_n and v_n', the corresponding version obtained from samples with underlying distributions μ_j'. In fact, arguing as above, we can write

$$v_n^2 = \inf_{\varphi \in \mathcal{G}}\left[\inf_{\pi \in \Pi(U_1, \ldots, U_J)} \int T(\varphi; i_1, \ldots, i_J) d\pi(i_1, \ldots, i_J)\right],$$

where $T(\varphi; i_1, \ldots, i_J) = \min_{y \in \mathbb{R}} \frac{1}{J}\sum_{j=1}^{J}(Z_{i_j,j}(\varphi) - y)^2$. We write $T'(\varphi; i_1, \ldots, i_J)$ for the same function computed on the $Z_{i,j}'(\varphi)$'s and set

$$\|\varphi'\|_\infty := \sup_{\substack{x \in (c;d) \\ \varphi \in \mathcal{G}}} |\varphi_j'(x)|.$$

Now, from the fact $(Z_{i,j}(\varphi) - Z_{i,j}'(\varphi))^2 \leqslant \|\varphi'\|_\infty^2 (X_{i,j} - X_{i,j}')^2$ we see that

$$|T(\varphi; i_1, \ldots, i_J)^{1/2} - T'(\varphi; i_1, \ldots, i_J)^{1/2}|^2 \leqslant \|\varphi'\|_\infty^2 \frac{1}{J}\sum_{j=1}^{J}(X_{i_j,j} - X_{i_j,j}')^2$$

and, as a consequence, that

$$\begin{aligned} &|V_2\left(\mu_1^n(\varphi), \ldots, \mu_J^n(\varphi)\right) - V_2\left({\mu_1'}^n(\varphi), \ldots, {\mu_J'}^n(\varphi)\right)|^2 \\ &\qquad \leqslant \quad \frac{1}{J}\sum_{j=1}^{J}\sum_{i_j=1}^{n_j}\frac{1}{n_j}\|\varphi'\|_\infty^2 (X_{i_j,j} - X_{i_j,j}')^2 \end{aligned}$$

and then

$$(v_n - v_n')^2 \leqslant \|\varphi'\|_\infty^2 \frac{1}{J}\sum_{j=1}^{J}\left(\tfrac{1}{n_j}\textstyle\sum_{i=1}^{n_j}(X_{i,j} - X_{i,j}')^2\right).$$

If, as in the proof of Theorem 3.2, we assume that $(X_{i,j}, X_{i,j}')$, $i = 1, \ldots, n_j$ are i.i.d. copies of an optimal coupling for μ_j and μ_j', with different samples independent from each other we obtain that

$$\mathbb{E}\left[(v_n - v_n')^2\right] \leqslant \|\varphi'\|_\infty^2 \frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j, \mu_j').$$

$\square$

Proof of Corollary 4.3. In Theorem 4.2, take $\mu'_j = \mu_{n,j}$, and set $v^*_{m_n} := \inf_{\varphi\in\mathcal{G}} V_2(\mu^*_{m_n,1}(\varphi),\ldots,\mu^*_{m_n,J}(\varphi))$. Then, conditionally to the $X_{1,j},\ldots,X_{n,j}$, the result of Theorem 4.2 reads now

$$W_2^2(\mathcal{L}(v_{m_n}),\mathcal{L}(v^*_{m_n})) \leqslant \sup_{\varphi\in\mathcal{G}} \left\|\varphi'_j\right\|_\infty^2 \frac{1}{J}\sum_{j=1}^J W_2^2(\mu_j,\mu_{n,j}).$$

Now, let $v^2 := \inf_{\varphi\in\mathcal{G}} M(\varphi)$. Then,

$$\begin{aligned} W_2^2(\mathcal{L}(v_{m_n}),\mathcal{L}(v^*_{m_n})) &= W_2^2(\mathcal{L}(v_{m_n}-v),\mathcal{L}(v^*_{m_n}-v)) \\ &\leqslant \sup_{\varphi\in\mathcal{G}} \left\|\varphi'_j\right\|_\infty^2 \frac{1}{J}\sum_{j=1}^J W_2^2(\mu_j,\mu_{n,j}). \end{aligned} \tag{10}$$

Now, recall that in the proof of Theorem 4.1 one gets that $W_2^2(\mu_j,\mu_{n,j}) = O_{\mathbb{P}}(\frac{1}{\sqrt{n}})$. Then, using that $W_r(\mathcal{L}(aX),\mathcal{L}(aY)) = aW_r(\mathcal{L}(X),\mathcal{L}(Y))$ for $a>0$, (10) gives

$$\begin{aligned} &W_2^2\left(\mathcal{L}\left(\sqrt{m_n}\left(v_{m_n}-v\right)\right),\mathcal{L}\left(\sqrt{m_n}\left(v^*_{m_n}-v\right)\right)\right) \\ &\qquad\leqslant \frac{m_n}{\sqrt{n}}\sup_{\varphi\in\mathcal{G}} \left\|\varphi'_j\right\|_\infty^2 \frac{1}{J}\sum_{j=1}^J nW_2^2(\mu_j,\mu_{n,j}) \to 0 \end{aligned} \tag{11}$$

Moreover, under Assumptions $\mathcal{A}$**1** to $\mathcal{A}$**6**, Theorem 4.1 gives $\sqrt{m_n}\left(v^2_{m_n}-v^2\right) \rightharpoonup \gamma$. If $v>0$, the classical Delta Method (see for instance in Van der Vaart (2000) p.25) gives

$$\sqrt{m_n}\left(v_{m_n}-v\right) \rightharpoonup \frac{1}{2v}\gamma.$$

Hence (11) enables to say that that

$$\sqrt{m_n}\left(v^*_{m_n}-v\right) \rightharpoonup \frac{1}{2v}\gamma.$$

Applying again a Delta Method leads to

$$\sqrt{m_n}\left((v^*)^2_{m_n}-v^2\right) = \sqrt{m_n}\left(\inf_{\mathcal{G}} U^*_{m_n} - \inf_{\mathcal{G}} U\right) \rightharpoonup \gamma.$$

(7) is obtained by using Glivenko Cantelli Theorem and convergence of the empirical quantiles.

$\square$

6.2. Proofs for the deformation model

We provide here proofs of the main results in Sections 3 and 4. Our approach relies on the consideration of quantile processes, namely,

$$\rho_{n,j}(t) = \sqrt{n} f_j(F_j^{-1}(t))(F_{n,j}^{-1}(t) - F_j^{-1}(t)), \quad 0<j<1,\ j=1,\ldots,J,$$

and on strong approximations of quantile processes, as in the following result that we adapt from Csörgő and Horváth (1993) (Theorem 2.1, p. 381 there).

Theorem 6.1. *Under $\mathcal{A}$**2**, there exist, on a rich enough probability space, inependent versions of $\rho_{n,j}$ and independent families of Brownian bridges $\{B_{n,j}\}_{n=1}\infty$, $j = 1, \ldots, J$ satisfying*

$$n^{1/2-\nu} \sup_{1/n \leqslant t \leqslant 1-1/n} \frac{|\rho_{n,j}(t) - B_{n,j}(t)|}{(t(1-t))^{\nu}} = \begin{cases} O_p(\log(n)) \text{ if } \nu = 0 \\ O_p(1) \text{ if } 0 < \nu \leqslant 1/2 \end{cases}$$

We will make frequent use in this section of the following technical Lemma which generalizes a result in Álvarez-Esteban et al. (2008).

Lemma 6.2. *Under Assumption $\mathcal{A}$**6***

i) $\sup_{h \in \mathcal{G}_j} \sqrt{n} \int_0^{\frac{1}{n}} (h(F_j^{-1}(t)))^2 dt \to 0$, $\sup_{h \in \mathcal{G}_j} \sqrt{n} \int_{1-\frac{1}{n}}^1 (h(F_j^{-1}(t)))^2 dt \to 0$.

ii) $\sup_{h \in \mathcal{G}_j} \sqrt{n} \int_0^{\frac{1}{n}} (h(F_{n,j}^{-1}(t)))^2 dt \to 0$, $\sup_{h \in \mathcal{G}_j} \sqrt{n} \int_{1-\frac{1}{n}}^1 (h(F_{n,j}^{-1}(t)))^2 dt \to 0$ *in probability.*

*iii) If moreover $\mathcal{A}$**3** holds*

$$\forall k, j \int_0^1 \frac{\sqrt{t(1-t)}}{f_k\left(F_k^{-1}(t)\right)} \sup_{\varphi \in \mathcal{G}} \left|\varphi_j\left(F_j^{-1}(t)\right) - F_B^{-1}(\varphi)(t)\right| dt < \infty \qquad (12)$$

iv) In the parametric case, under Assumptions **A3**, **A6** *and if $\forall k$, F_k is C^1 with $F_k' = f_k > 0$ on (c_k, d_k)*

$$\forall k, j \int_0^1 \frac{\sqrt{t(1-t)}}{f_k\left(F_k^{-1}(t)\right)} \sup_{\theta \in \Theta} \left|\varphi_{\theta_j}^{-1}\left(F_j^{-1}(t)\right) - F_B^{-1}(\theta)(t)\right| dt < \infty \qquad (13)$$

Our next proof is inspired by Álvarez-Esteban et al. (2008). The main part concerns the study of $\sqrt{n}U_n(\varphi)$ uniformly in φ in probability by using strong approximations of the quantile process with Brownian bridges.

Proof of Theorem 4.1. We will work with the versions of $\rho_{n,j}$ and $B_{n,j}$ given by Theorem 6.1. We show first that

$$\sup_{\varphi \in \mathcal{G}} \left|\sqrt{n}\left(U_n(\varphi) - U(\varphi)\right) - \frac{1}{J}\sum_{j=1}^J S_{n,j}(\varphi)\right| \to 0 \text{ in probability} \qquad (14)$$

with $S_{n,j}(\varphi) = 2\int_0^1 \varphi_j' \circ F_j^{-1}(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)) \frac{B_{n,j}}{f_j \circ F_j^{-1}}$. To check this we note that the fact that $\frac{1}{J}\sum_{j=1}^J \varphi_j \circ F_j^{-1} = F_B^{-1}(\varphi)$ and simple algebra yield $\sqrt{n}(U_n(\varphi) - U(\varphi)) = \frac{2}{J}\sum_{j=1}^J \tilde{S}_{n,j} + \frac{1}{J}\sum_{j=1}^J \tilde{R}_{n,j}$ with

$$\tilde{S}_{n,j} = \sqrt{n}\int_0^1 (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)),$$

$$\tilde{R}_{n,j} = \sqrt{n}\int_0^1 [(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}) - (F_{n,B}^{-1}(\varphi) - F_B^{-1}(\varphi))]^2.$$

From the elementary inequality $(a_1 + \cdots + a_J)^2 \leq Ja_1^2 + \cdots + Ja_J^2$ we get that

$$\frac{1}{J}\sum_{j=1}^J \tilde{R}_{n,j} \leq \frac{4\sqrt{n}}{J}\sum_{j=1}^J \int_0^1 (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})^2$$

Now, for every $t \in (0,1)$ we have

$$\varphi_j \circ F_{n,j}^{-1}(t) - \varphi_j \circ F_j^{-1}(t) = \varphi_j'(K_{n,\varphi_j}(t))(F_{n,j}^{-1}(t) - F_j^{-1}(t)) \tag{15}$$

for some $K_{n,\varphi_j}(t)$ between $F_{n,j}^{-1}(t)$ and $F^{-1}(t)$. Assumption $\mathcal{A}$**4** implies $C_j := \sup_{\varphi_j \in \mathcal{G}_j, x \in (c_j, d_j)} |\varphi_j'(x)| < \infty$. Hence, we have

$$\int_0^1 (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})^2 \leq C_j^2 \int_0^1 (F_{n,j}^{-1} - F_j^{-1})^2.$$

Now we can use $\mathcal{A}$**5** and argue as in the proof of Theorem 2 in Álvarez-Esteban et al. (2008) to conclude that $\sqrt{n}\int_0^1 (F_{n,j}^{-1} - F_j^{-1})^2 \to 0$ in probability and, as a consequence, that

$$\sup_{\varphi \in \mathcal{G}} \left| \sqrt{n}\left(U_n(\varphi) - U(\varphi)\right) - \frac{1}{J}\sum_{j=1}^J \tilde{S}_{n,j}(\varphi) \right| \to 0 \text{ in probability.} \tag{16}$$

On the other hand, the Cauchy-Schwarz's inequality shows that

$$\begin{aligned} n\Big(\int_0^{\frac{1}{n}} (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))\Big)^2 & \\ \leq \quad \sqrt{n}\int_0^{\frac{1}{n}} (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})^2 \sqrt{n}\int_0^{\frac{1}{n}} (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))^2 & \end{aligned}$$

and using i) and ii) of Lemma 6.2, the two factors converge to zero uniformly in φ. A similar argument works for the upper tail and allows to conclude that we can replace in (16) $\tilde{S}_{n,j}(\varphi)$ with $\tilde{\tilde{S}}_{n,j}(\varphi) := 2\sqrt{n}\int_{\frac{1}{n}}^{1-\frac{1}{n}} (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))$. Moreover,

$$\begin{aligned} \sup_{\varphi \in \mathcal{G}} \Big| \int_0^{\frac{1}{n}} \varphi_j' \circ F_j^{-1} \frac{B_{n,j}}{f_j \circ F_j^{-1}} (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)) \Big| & \\ \leq \quad C_j \int_0^{\frac{1}{n}} \Big| \frac{B_{n,j}}{f_j \circ F_j^{-1}} \Big| \sup_{\varphi \in \mathcal{G}} \big| (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)) \big| & \end{aligned}$$

and by iii) of Lemma 6.2 and Cauchy-Schwarz's inequality

$$\begin{aligned} \mathbb{E}\Big[\int_0^{\frac{1}{n}} \Big| \frac{B_{n,j}}{f_j \circ F_j^{-1}} \Big| \sup_{\varphi \in \mathcal{G}} \big| (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)) \big| \Big] & \\ \leq \quad \int_0^{\frac{1}{n}} \frac{\sqrt{t(1-t)}}{f_j(F_j^{-1}(t))} \sup_{\varphi \in \mathcal{G}} \big| \varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t) \big| dt \to 0. & \end{aligned}$$

Hence, $\sup_{\varphi\in\mathcal{G}}\Big|\int_0^{\frac{1}{n}}\varphi_j'\circ F_j^{-1}\frac{B_{n,j}}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big|\to 0$ in probability and similarly for the right tail. Thus (recall (15)), to prove (14) it suffices to show that

$$\sup_{\varphi\in\mathcal{G}}\Big|\int_{\frac{1}{n}}^{1-\frac{1}{n}}\varphi_j'(F_j^{-1}(t))\frac{B_{n,j}(t)}{f_j(F_j^{-1}(t))}(\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t))dt \tag{17}$$
$$-\quad\int_{\frac{1}{n}}^{1-\frac{1}{n}}\varphi_j'(K_{n,\varphi_j}(t))\frac{\rho_{n,j}(t)}{f_j(F_j^{-1}(t))}(\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t))dt\Big|\to 0$$

in probability. To check it we take $\nu\in(0,1/2)$ and use Theorem 6.1 to get

$$\begin{aligned}\int_{\frac{1}{n}}^{1-\frac{1}{n}}&\frac{|\rho_{n,j}(t)-B_{n,j}(t)|}{f_j(F_j^{-1}(t))}\sup_{\varphi\in\mathcal{G}}\big|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)\big|dt\\&\leqslant n^{\nu-\frac{1}{2}}O_P(1)\int_{\frac{1}{n}}^{1-\frac{1}{n}}\frac{(t(1-t))^{\nu}}{f_k(F_k^{-1}(t))}\sup_{\varphi\in\mathcal{G}}\big|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)\big|dt\to 0\end{aligned}\tag{18}$$

in probability (using dominated convergence and iii) of Lemma 6.2).

We observe next that, for all $t\in(0,1)$, $\sup_{\varphi_j\in\mathcal{G}_j}|K_{n,\varphi_j}(t)-F_j^{-1}(t)|\to 0$ almost surely, since $K_{n,\varphi_j}(t)$ lies between $F_{n,j}^{-1}(t)$ and $F_j^{-1}(t)$. Therefore, using Assumption $\mathcal{A}4$ we see that $\sup_{\varphi_j\in\mathcal{G}_j}|\varphi_j'(K_{n,\varphi_j}(t))-\varphi_j'(F_j^{-1}(t)|\to 0$ almost surely while, on the other hand $\sup_{\varphi_j\in\mathcal{G}_j}|\varphi_j'(K_{n,\varphi_j}(t))-\varphi_j'(F_j^{-1}(t))|\leq 2C_j$. But then, by dominated convergence we get that

$$\mathbb{E}\Big[\sup_{\varphi_j\in\mathcal{G}_j}|\varphi_j'(K_{n,\varphi_j}(t))-\varphi_j'(F_j^{-1}(t))|^2\Big]\to 0.$$

Since by iii) of Lemma 6.2 we have that $t\mapsto\frac{\sqrt{t(1-t)}}{f_j(F_j^{-1}(t))}\sup_{\varphi\in\mathcal{G}}|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)|$ is integrable we conclude that

$$\mathbb{E}\sup_{\varphi\in\mathcal{G}}\int_{\frac{1}{n}}^{1-\frac{1}{n}}|\varphi_j'(K_{n,\varphi_j}(t))-\varphi_j'(F_j^{-1}(t))|\frac{|B_{n,j}(t)|}{f_j(F_j^{-1}(t))}|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)|dt$$

tends to 0 as $n\to\infty$ and, consequently,

$$\sup_{\varphi\in\mathcal{G}}\int_{\frac{1}{n}}^{1-\frac{1}{n}}|\varphi_j'(K_{n,\varphi_j}(t))-\varphi_j'(F_j^{-1}(t))|\frac{|B_{n,j}(t)|}{f_j(F_j^{-1}(t))}|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)|dt$$

vanishes in probability. Combining this fact with (18) we prove (17) and, as a consequence, (14).

Observe now that for all $n\in\mathbb{N}$, $(S_{n,j}(\varphi))_{1\leqslant j\leqslant J}$ has the same law as $(S_j(\varphi))_{1\leqslant j\leqslant J}$ with

$$S_j(\varphi)=2\int_0^1\varphi_j'\circ F_j^{-1}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\frac{B_j}{f_j\circ F_j^{-1}}$$

and $(B_j)_{1\leqslant j\leqslant J}$ independent standard Brownian bridges. Set $S=\frac{1}{J}\sum_{j=1}^{J}S_j$. Now, (14) implies that

$$\sqrt{n}\left(U^n\left(\cdot\right)-U\left(\cdot\right)\right)\rightharpoonup S\left(\cdot\right) \tag{19}$$

in the space $L^{\infty}\left(\mathcal{G}\right)$ (we denote by $\left\|\cdot\right\|_{\infty}$ the norm on this space). From Skohorod Theorem we know that there exists some probability space on which the convergence (19) holds almost surely. From now on, we place us on this space. Then, for $\varphi,\rho\in\mathcal{G}$

$$\begin{aligned}
|S_j(\varphi)-S_j(\rho)| &\leq 2\sup_{(c_j,d_j)}|\varphi_j'-\rho_j'|\Big|\int_0^1\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big|\\
&+2\Big|\int_0^1\frac{B_j}{f_j\circ F_j^{-1}}\rho_j'\circ F_j^{-1}(\varphi_j\circ F_j^{-1}-\rho_j\circ F_j^{-1})\Big|\\
&\leq 2\sup_{(c_j,d_j)}|\varphi_j'-\rho_j'|\sup_{\varphi\in\mathcal{G}}\Big|\int_0^1\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big|\\
&+2\sup_{(c_j,d_j)}|\rho_j'|\Big(\textstyle\int_0^1\big|\frac{B_j}{f_j\circ F_j^{-1}}\big|^q\Big)^{1/q}\Big(\textstyle\int_0^1|\varphi_j\circ F_j^{-1}-\rho_j\circ F_j^{-1}|^{p_0}\Big)^{1/p_0}
\end{aligned}$$

But using iii) of Lemma 6.2

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\varphi\in\mathcal{G}}\Big|\int_0^1\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big|\Big]\\
&\qquad\leqslant\int_0^1\frac{\sqrt{t(1-t)}}{f_k\left(F_k^{-1}(t)\right)}\sup_{\varphi\in\mathcal{G}}\left|\varphi_j\left(F_j^{-1}\left(t\right)\right)-F_B^{-1}\left(\varphi\right)\left(t\right)\right|dt<\infty
\end{aligned}$$

Hence, almost surely, $\sup_{\varphi\in\mathcal{G}}\left|\int_0^1\frac{B_j}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\right|<\infty$. Furthermore, from Assumption $\mathcal{A}\mathbf{3}$, we get that a.s.

$$\int_0^1\left(\tfrac{B_j}{f_j\circ F_j^{-1}}\right)^q<\infty$$

and thus, for some random variable T a.s. finite , and φ, $\rho\in\mathcal{G}$, we get

$$\left|S_j\left(\varphi\right)-S_j\left(\rho\right)\right|\leqslant T\left\|\varphi-\rho\right\|_{\mathcal{G}}.$$

Thus, we deduce that $(S_j)_{1\leqslant j\leqslant J}$ are almost surely continuous functions on $\mathcal{G}$, endowed with the norm $\left\|\cdot\right\|_{\mathcal{G}}$.

Observe now that

$$\sqrt{n}\Big(\inf_{\mathcal{G}}U_n-\inf_{\mathcal{G}}U\Big)\leq\sqrt{n}\inf_{\Gamma}U_n-\sqrt{n}\inf_{\Gamma}U=\inf_{\Gamma}\sqrt{n}\left(U_n-U\right). \tag{20}$$

On the other hand, if we consider the (a.s.) compact set $\Gamma_n=\{\varphi\in\mathcal{G}:U\left(\varphi\right)\leqslant\inf_{\mathcal{G}}U+\frac{2}{\sqrt{n}}\left\|\sqrt{n}\left(U_n-U\right)\right\|_{\infty}\}$, then, if $\varphi\notin\Gamma^n$,

$$U_n\left(\varphi\right)\geqslant\inf_{\mathcal{G}}U+2\left\|\left(U_n-U\right)\right\|_{\infty}-\left\|\left(U_n-U\right)\right\|_{\infty},$$

which implies

$$U_n(\varphi) \geqslant \inf_{\mathcal{G}} U + \|(U_n - U)\|_\infty,$$

while if $\varphi \in \Gamma$, then,

$$U_n(\varphi) = \inf_{\mathcal{G}} U + U^n(\varphi) - U(\varphi) \leqslant \inf_{\mathcal{G}} U + \|(U_n - U)\|_\infty.$$

Thus, necessarily, $\inf_{\mathcal{G}} U_n = \inf_{\Gamma_n} U_n = \inf_{\Gamma_n}(U_n - U + U) \geq \inf_{\Gamma_n}(U_n - U) + \inf_{\Gamma_n} U = \inf_{\Gamma_n}(U_n - U) + \inf_{\Gamma} U$. Together with (20) this entails

$$\inf_{\Gamma_n} \sqrt{n}(U_n - U) \leqslant \sqrt{n}\big(\inf_{\mathcal{G}} U_n - \inf_{\mathcal{G}} U\big) \leqslant \inf_{\Gamma} \sqrt{n}(U_n - U). \tag{21}$$

Note that for the versions that we are considering $\|\sqrt{n}(U_n - U) - S\|_\infty \to 0$ a.s.. In particular, this implies that $\inf_\Gamma \sqrt{n}(U_n - U) \to \inf_\Gamma S$ a.s.. Hence, the proof will be complete if we show that a.s.

$$\inf_{\Gamma_n} \sqrt{n}(U_n - U) \to \inf_{\Gamma} S. \tag{22}$$

To check this last point, consider a sequence $\varphi_n \in \Gamma_n$ such that $\sqrt{n}(U_n(\varphi_n) - U(\varphi_n)) \leq \inf_{\Gamma_n} \sqrt{n}(U_n - U) + \frac{1}{n}$. By compactness of $\mathcal{G}$, taking subsequences if necessary, $\varphi_n \to \varphi_0$ for some $\mathcal{G}$. Continuity of U yields $U(\varphi_n) \to U(\varphi_0)$ and as a consequence, that $U(\varphi_0) \leq \inf_{\mathcal{G}} U$, that is, $\varphi_0 \in \Gamma$ a.s.. Furthermore,

$$\begin{aligned}&\big|\sqrt{n}(U_n - U)(\varphi_n) - S(\varphi_0)\big| \\ &\qquad\leqslant \big\|\sqrt{n}(U_n - U) - S\big\|_\infty + |S(\varphi_n) - S(\varphi_0)| \to 0.\end{aligned}$$

This shows that

$$\liminf \inf_{\Gamma_n} \sqrt{n}(U_n - U) \geq S(\varphi_0) \geqslant \inf_{\Gamma} S \tag{23}$$

and yields (22). This completes the proof.

□

Proof of Proposition 5.2. We denote by ∂_j the derivative operator w.r.t. θ_j, $1 \leqslant j \leqslant J$ and $\partial_{j,k}$ for second order partial derivatives. We note that $\mathcal{H}$ entails that the empirical d.f. on the j-th sample, $F_{n,j}(t)$, satisfies $F_{n,j}(t) = G_{n,j}(\varphi_{\theta_j^*}(t))$ with $G_{n,j}$ the empirical d.f. on the $\varepsilon_{i,j}$'s (which are i.i.d. μ, with d.f. G). We write now $\rho_{n,j}$ for the quantile process based on the $\varepsilon_{i,j}$'s. We write $B_{n,j}$ for independent Brownian bridges as given by Theorem 6.1 (observe that (**A2**) grants the existence of such $B_{n,j}$'s).

Assumption **TCL** implies that $\partial\varphi_{\theta_j^*} \in L^2(X_j)$. Moreover, with Assumptions **A8**, **A9** and compactness of Θ, we deduce that $\sup_{\lambda\in\Lambda} \partial\varphi_\lambda \in L^2(X_j)$. On the other hand, since ε has a moment of order $r > 4$, arguing as in the proof of point 3 in Lemma 6.2 we have that

$$\int_0^1 \frac{\sqrt{t(1-t)}}{g(G^{-1}(t))} dt < \infty. \tag{24}$$

From **A8** and **A9** we have that U_n is a C^2 function and derivatives can be omputed by differentiation under the integral sign. This implies that

$$\partial_j U_n(\theta) = \frac{2}{J}\int_0^1 \partial\varphi_{\theta_j}(F_{n,j}^{-1}(t))\Big(\varphi_{\theta_j}(F_{n,j}^{-1}(t)) - \frac{1}{J}\sum_{k=1}^J \varphi_{\theta_k}(F_{n,k}^{-1}(t))\Big)dt,$$

$$\partial^2_{p,q}U_n(\theta) = -\frac{2}{J^2}\int_0^1 \partial\varphi_{\theta_p}(F_{n,p}^{-1}(t))\partial\varphi_{\theta_q}(F_{n,q}^{-1}(t))dt, \quad p \neq q \qquad (25)$$

and

$$\begin{aligned}\partial^2_{p,p}U_n(\theta) \quad &= \quad \frac{2}{J}\int_0^1 \partial^2\varphi_{\theta_p}((F_{n,p})^{-1}(t))\Big(\varphi_{\theta_p}(F_{n,p}^{-1}(t)) - \frac{1}{J}\sum_{k=1}^J \varphi_{\theta_k}(F_{n,k}^{-1}(t))\Big)\\ &\quad + \frac{2(J-1)}{J^2}\int_0^1 (\partial\varphi_{\theta_p}(F_{n,p}^{-1}(t))^2 dt.\end{aligned}$$

Similar expressions are obtained for the derivatives of $U(\theta)$ (replacing everywhere $F_{n,j}^{-1}$ with $F_j^{-1} = \varphi_{\theta_j^*}^{-1} \circ G^{-1}$). We write $DU_n(\theta) = (\partial_j U_n(\theta))_{1\le j\le J}$, $DU(\theta) = (\partial_j U(\theta))_{1\le j\le J}$ for the gradients and $\Phi_n(\theta) = [\partial^2_{p,q}U_n(\theta)]_{1\le p,q\le J}$, $\Phi(\theta) = [\partial^2_{p,q}U(\theta)]_{1\le p,q\le J}$for the Hessians of U_n and U. Note that $\Phi^* = \Phi(\theta^*)$ is assumed to be invertible.

Recalling that $R_j = \partial\varphi_{\theta_j^\star} \circ \varphi_{\theta_j^\star}^{-1}$, from the fact $DU(\theta^*) = 0$ we see that

$$\sqrt{n}\partial_j U_n(\theta^\star) = \frac{2}{J}\int_0^1 R_j(G_{n,j}^{-1}(t))\frac{\rho_{n,j}(t) - \frac{1}{J}\sum_{k=1}^J \rho_{n,k}(t)}{g(G^{-1}(t))}dt. \qquad (26)$$

Now, using Assumption **TCL** and arguing as in the proof of Theorem 4.1 we conclude that

$$\left|\int_0^1 R_j(G_{n,j}^{-1}(t))\frac{\rho_{n,k}(t)}{g\left(G^{-1}(t)\right)}dt - \int_0^1 R_j\left(G^{-1}(t)\right)\frac{B_{n,k}(t)}{g\left(G^{-1}(t)\right)}dt\right| \to 0$$

in probability and, consequently,

$$\left|\sqrt{n}\partial_j U_n(\theta^\star) - \frac{2}{J}\int_0^1 R_j\left(G^{-1}(t)\right)\frac{B_{n,j}(t) - \frac{1}{J}\sum_{k=1}^J B_{n,k}(t)}{g\left(G^{-1}(t)\right)}dt\right| \to 0 \qquad (27)$$

in probability.

A Taylor expansion of $\partial_j U_n$ around θ^* shows that for some $\tilde{\theta}_j^n$ between $\hat{\theta}^n$ and $\theta^\star$ we have

$$\partial_j U_n(\hat{\theta}^n) = \partial_j U^n(\theta^\star) + (\partial^2_{1j}U_n(\tilde{\theta}_j^n), \dots, \partial^2_{Jj}U_n(\tilde{\theta}_j^n)) \cdot (\hat{\theta}^n - \theta^\star)$$

and because $\hat{\theta}^n$ is a zero of DU_n, we obtain

$$-\partial_j U^n(\theta^\star) = (\partial^2_{1j}U_n(\tilde{\theta}_j^n), \dots, \partial^2_{Jj}U_n(\tilde{\theta}_j^n)) \cdot (\hat{\theta}^n - \theta^\star).$$

Writing $\tilde{\Phi}_n$ for the $(J-1)\times(J-1)$ matrix whose $J-1$-th row equals $(\partial^2_{1j}U_n(\tilde{\theta}^n_j),\ldots,\partial^2_{Jj}U_n(\tilde{\theta}^n_j))$, $j=2,\ldots,J$, we can rewrite the last expansion as

$$-\sqrt{n}DU_n(\theta^*)=\tilde{\Phi}_n\sqrt{n}(\hat{\theta}^n-\theta^\star). \tag{28}$$

We show next that $\tilde{\Phi}_n\to\Phi^*=\Phi(\theta^*)$ in probability. Recalling (25), we consider first $\int_0^1(\partial\varphi_{\tilde{\theta}^n_p}(F^{-1}_{n,p}(t)))^2dt$. We have

$$\begin{aligned}
&\Big(\int_0^1(\partial\varphi_{\tilde{\theta}^n_p}(F^{-1}_{n,p}(t))-\partial\varphi_{\theta^*_p}(F^{-1}_p(t)))^2dt\Big)^{1/2}\\
&\quad\leq\ \Big(\int_0^1(\partial\varphi_{\tilde{\theta}^n_p}(F^{-1}_{n,p}(t))-\partial\varphi_{\theta^*_p}(F^{-1}_{n,p}(t)))^2dt\Big)^{1/2}\\
&\qquad+\Big(\int_0^1(\partial\varphi_{\theta^*_p}(F^{-1}_{n,p}(t))-\partial\varphi_{\theta^*_p}(F^{-1}_p(t)))^2dt\Big)^{1/2}\\
&\quad\leq\ \Big(\int_0^1\sup_{\lambda\in\Lambda}\left|\partial^2\varphi_\lambda(F^{-1}_{n,p}(t))\right|^2dt\Big)^{1/2}|\tilde{\theta}^n_p-\theta^*_p|\\
&\qquad+\Big(\int_0^1(R_p(G^{-1}_{n,p}(t))-R_p(G^{-1}_p(t)))^2dt\Big)^{1/2}\to 0
\end{aligned}$$

in probability, where we have used assumptions **A9**, **TCL** and Proposition 5.1. A similar argument shows that $\int_0^1(\varphi_{\tilde{\theta}^n_p}(F^{-1}_{n,p}(t))-\varphi_{\theta^*_p}(F^{-1}_p(t)))^2dt$ in probability. As a consequence, we conclude

$$\tilde{\Phi}_n\to\Phi^*,\quad \text{in probability.} \tag{29}$$

Now, (28), (27) (29) together with Slutsky's Theorem complete the proof.

$\square$

Proof of Theorem 5.3. We consider the same notation and setup as in the proof of Proposition 5.2. Since $DU_n(\hat{\theta}^n)=0$, a Taylor expansion around $\hat{\theta}^n$ shows that

$$nU_n(\theta^*)-nU_n(\hat{\theta}^n)=\frac{1}{2}(\sqrt{n}(\hat{\theta}^n-\theta^*))'\Phi(\tilde{\theta}_n)(\sqrt{n}(\hat{\theta}^n-\theta^*)) \tag{30}$$

for some $\tilde{\theta}_n$ between $\hat{\theta}^n$ and θ^*. Arguing as in the proof of Proposition 5.2 we see that $\Phi(\tilde{\theta}_n)\to\Phi^*$ in probability. Hence, to complete the proof if suffices to show that

$$nU_n(\theta^*)-\frac{1}{J}\sum_{j=1}^k\int_0^1\frac{\big(B_{n,j}(t)-\frac{1}{J}\sum_{k=1}^J B_{n,k}(t)\big)^2}{g(G^{-1}(t))^2}dt\to 0$$

in probability. Since

$$nU_n(\theta^*)=\frac{1}{J}\sum_{j=1}^k\int_0^1\frac{\big(\rho_{n,j}(t)-\frac{1}{J}\sum_{k=1}^J \rho_{n,k}(t)\big)^2}{g(G^{-1}(t))^2}dt,$$

this amounts to proving that

$$\int_0^1 \frac{\big(\rho_{n,j}(t)-B_{n,j}(t)\big)^2}{g(G^{-1}(t))^2}dt \to 0$$

in probability.

Taking $\nu \in (0,\frac{1}{2})$ in Theorem 6.1 we see that

$$\int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{\big(\rho_{n,j}(t)-B_{n,j}(t))\big)^2}{g(G^{-1}(t))^2}dt \le O_P(1)\frac{1}{n^{1-2\nu}}\int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{(t(1-t))^{2\nu}}{g(G^{-1}(t))^2} \to 0,$$

using condition (**A10**) and dominated convergence. From (**A10**) we also see that $\int_{1-\frac{1}{n}}^1 \frac{B_{n,j}(t)^2}{g(G^{-1}(t))^2}dt \to 0$ in probability. Condition (**A10**) implies also that $\int_{1-\frac{1}{n}}^1 \frac{\rho_{n,j}(t)^2}{g(G^{-1}(t))^2}dt \to 0$ in probability, see Samworth and Johnson (2004). Similar considerations apply to the left tail and complete the proof.

□

References

Agueh, M. and Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182 (2012e:49090)

Agulló-Antolín, M., Cuesta-Albertos, J. A., Lescornel, H. and Loubes, J.-M. (2015). A parametric registration model for warped distributions with Wasserstein's distance. *J. Multivariate Anal.* **135** 117–130. MR3306430

Álvarez-Esteban, P. C., del Barrio, E., Cuesta-Albertos, J. A. and Matrán, C. (2008). Trimmed comparison of distributions. *J. Amer. Statist. Assoc.* **103** 697–704. MR2435470 (2009i:62036)

Boissard, E., Le Gouic, T. and Loubes, J.-M. (2015). Distribution's template estimate with Wasserstein metrics. *Bernoulli* **21** 740–759. MR3338645

Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2014). Monge-Kantorovich Depth, Quantiles, Ranks, and Signs. *ArXiv e-prints*.

Collier, O. and Dalalyan, A. S. (2015). Curve registration by nonparametric goodness-of-fit testing. *J. Statist. Plann. Inference* **162** 20–42. MR3323102

Csörgő, M. (1983). *Quantile processes with statistical applications*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **42**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR745130 (86g:60045)

Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons Ltd., Chichester. With a foreword by David Kendall. MR1215046 (94c:60060)

Czado, C. and Munk, A. (1998). Assessing the similarity of distributions—finite sample performance of the empirical Mallows distance. *J. Statist. Comput. Simulation* **60** 319–346. MR1704844 del Barrio, E., Deheuvels, P. and van de Geer, S. (2007). *Lectures on Empirical Processes: Theory and Statistical Applications*. European Mathematical Society.

Freitag, G. and Munk, A. (2005). On Hadamard differentiability in k-sample semiparametric models—with applications to the assessment of structural relationships. *Journal of Multivariate Analysis* **94** 123–158.

Gallón, S., Loubes, J.-M. and Maza, E. (2013). Statistical properties of the quantile normalization method for density curve alignment. *Mathematical Biosciences* **242** 129–142.

Gamboa, F., Loubes, J.-M. and Maza, E. (2007). Semi-parametric Estimation of Shifts. *Electronic Journal of Statistics* **1** 616–640.

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 223–241. MR1625620 (99d:62052)

Rachev, S. T. (1984). The Monge-Kantorovich problem on mass transfer and its applications in stochastics. *Teor. Veroyatnost. i Primenen.* **29** 625–653. MR773434 (86m:60026)

Rajput, B. S. (1972). Gaussian measures on L_p spaces, $1 \leq p < \infty$. *J. Multivariate Anal.* **2** 382–403. MR0345157 (49 ##9896)

Samworth, R. and Johnson, O. (2004). Convergence of the empirical process in Mallows distance, with an application to bootstrap performance. *ArXiv e-prints*.

Van der Vaart, A. W. (2000). *Asymptotic statistics* **3**. Cambridge Univ Pr.

Villani, C. (2009). *Optimal transport: old and new* **338**. Springer Verlag.