Journal of Multivariate Analysis 151 (2016) 90–109

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Limit laws of the empirical Wasserstein distance: Gaussian distributions

Thomas Rippl [a,*], Axel Munk [a,b], Anja Sturm [a]

[a] *Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Göttingen, Germany*

[b] *Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

ARTICLE INFO

Article history:
Received 17 July 2015
Available online 19 July 2016

AMS 2010 subject classifications:
primary 62E20
60F05
secondary 58C20
90B06

Keywords:
Mallow's metric
Transport metric
Delta method
Limit theorem
Goodness-of-fit
Fréchet derivative
Resolvent operator
Bootstrap
Elliptically symmetric distribution

ABSTRACT

We derive central limit theorems for the Wasserstein distance between the empirical distributions of Gaussian samples. The cases are distinguished whether the underlying laws are the same or different. Results are based on the (quadratic) Fréchet differentiability of the Wasserstein distance in the gaussian case. Extensions to elliptically symmetric distributions are discussed as well as several applications such as bootstrap and statistical testing.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Let $\mathbb{P}, \mathbb{Q}$ be in $\mathcal{M}_1(\mathbb{R}^d)$, the probability measures on $\mathbb{R}^d$. Consider $\pi_i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, x = (x_1, x_2) \mapsto x_i, i = 1, 2$, the projections on the first or the second d-dimensional vector, and define

$$\Pi(\mathbb{P}, \mathbb{Q}) = \{\mu \in \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d) : \mu \circ \pi_1^{-1} = \mathbb{P}, \mu \circ \pi_2^{-1} = \mathbb{Q}\}$$

as the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathbb{P}$ and $\mathbb{Q}$. Then for $p \geq 1$ we define the *p-Wasserstein distance* as

$$\mathcal{W}_p(\mathbb{P}, \mathbb{Q}) := \inf_{\mu \in \Pi(\mathbb{P}, \mathbb{Q})} \left(\int_{\mathbb{R}^{2d}} \|x - y\|^p \, \mu(dx, dy) \right)^{1/p}. \tag{1}$$

There is a variety of interpretations and equivalent definitions of $\mathcal{W}_p$, for example as a mass transport problem; we refer the reader for extensive overviews to Villani [46] and Rachev and Rüschendorf [36].

* Corresponding author.
E-mail addresses: trippl@uni-goettingen.de (T. Rippl), munk@math.uni-goettingen.de (A. Munk), asturm@math.uni-goettingen.de (A. Sturm).

http://dx.doi.org/10.1016/j.jmva.2016.06.005
0047-259X/© 2016 Elsevier Inc. All rights reserved.

In this paper we are concerned with the *statistical* task of estimating $\mathcal{W}_p(\mathbb{P},\mathbb{Q})$ from given data $X_1,\dots,X_n \sim \mathbb{P}$ i.i.d. (and possibly also from data $Y_1,\dots,Y_m \sim \mathbb{Q}$ i.i.d.) and with the investigation of certain characteristics of this estimate which are relevant for inferential purposes. Replacing $\mathbb{P}$ by the empirical measure $\mathbb{P}_n$ associated with $X_1,\dots,X_n$ yields the empirical Wasserstein distance $\widehat{\mathcal{W}}_{p,n} := \widehat{\mathcal{W}_p}(\mathbb{P}_n,\mathbb{Q})$ which provides a natural estimate of $\mathcal{W}_p(\mathbb{P},\mathbb{Q})$ for a given $\mathbb{Q}$. Similarly, define $\widehat{\mathcal{W}}_{p,n,m} := \widehat{\mathcal{W}}_p(\mathbb{P}_n,\mathbb{Q}_m)$ in the two sample case. For inferential purposes (e.g., testing or confidence intervals for $\mathcal{W}_p(\mathbb{P},\mathbb{Q})$) it is of particular relevance to investigate the (asymptotic) distribution of the empirical Wasserstein distance.

This is meanwhile well understood for measures $\mathbb{P},\mathbb{Q}$ on the real line $\mathbb{R}$ as in this case an explicit representation of the Wasserstein distance (and its empirical counterpart) exists (see, e.g., [22,28,30,43,32,33])

$$\mathcal{W}_p^p(\mathbb{P},\mathbb{Q}) = \int_{[0,1]} |F^{-1}(t) - G^{-1}(t)|^p \, dt. \tag{2}$$

Here, $F(x) = \mathbb{P}((-\infty,x])$ and $G(x) = \mathbb{Q}((-\infty,x])$ for $x \in \mathbb{R}$ denote the c.d.f.s of $\mathbb{P}$ and $\mathbb{Q}$, respectively, and F^{-1} and G^{-1} its inverse quantile functions. Now, $\widehat{\mathcal{W}}_{p,n}$ is defined as in (2) with F^{-1} replaced by the empirical quantile function F_n^{-1}, and the representation (2) can be used to derive limit theorems based on the underlying quantile process $\sqrt{n}(F_n^{-1} - F^{-1})$. These results require a *scaling rate* $(a_n)_{n\in\mathbb{N}}$ such that the laws

$$a_n \left(\widehat{\mathcal{W}_p^p}(\mathbb{P}_n,\mathbb{Q}) - \mathcal{W}_p^p(\mathbb{P},\mathbb{Q}) + b_n \right), \quad \text{as } n \to \infty \tag{3}$$

(for some centering sequence $(b_n)_{n\in\mathbb{N}}$) converge weakly to a (non-degenerate) *limit distribution*. Depending on whether $F = G$ as well as on the tail behavior of the distributions F and G we find ourselves in different asymptotic regimes. Roughly speaking, when $F = G$ (i.e., $\mathbb{P} = \mathbb{Q}$, $\mathcal{W}_p(\mathbb{P},\mathbb{Q}) = 0$), $a_n = n$ is the proper scaling rate, i.e., the limit is of *second order* and given by a weighted sum of χ^2 laws (see e.g. [6,7]). In general, b_n depends on the tail behavior of F. In contrast, when $F \neq G$, i.e., $\mathcal{W}_p^p(\mathbb{P},\mathbb{Q}) > 0$ for $a_n = \sqrt{n}$, $b_n = 0$ the limit is of *first order* and $\sqrt{n}(\widehat{\mathcal{W}_p}(\mathbb{P}_n,\mathbb{Q}) - \mathcal{W}_p(\mathbb{P},\mathbb{Q}))$ is asymptotically normal (see [34,23]) under appropriate tail conditions. Various applications of these and related distributional results, e.g., for trimmed versions of the Wasserstein distance, include the comparison of distributions and goodness of fit testing [34,3,5,24], template registration (Section 4 in [11,1]), bioequivalence testing [23], atmospheric research [48], or large scale microscopy imaging [38].

In contrast to the real line ($d = 1$), up to now limiting results as in (3) remain elusive for $\mathbb{R}^d$, $d \geq 2$. However, see [2,18] for almost sure limit results and [21] for moment bounds on $\widehat{\mathcal{W}}_{p,n}$. Already the planar case $d = 2$ is remarkably challenging [2]. One difficulty is that no simple characterization as in (2) via the (empirical) c.d.f's exists anymore. In particular, the couplings for which the infimum in (1) is attained are much more involved, see, e.g., [31,37]. We will come back to this in the context of our subsequent results later on.

In this article we aim to shed some light on the case $d \geq 2$ by further restricting the possible measures $\mathbb{P},\mathbb{Q}$ to the Gaussians (and more generally to elliptical distributions). Here, a well known explicit representation of $\mathcal{W}_p(\mathbb{P},\mathbb{Q})$ can be used (see, e.g., [19,25,35]) which allows one to obtain explicit limit theorems again. The Gaussian case is of particular interest as it provides, as shown in [25], a universal lower bound for any pair $(\mathbb{P},\mathbb{Q})$ having the same moments (expectation and covariance) as the Gaussian law, see also [13].

Limit laws for the Gaussian Wasserstein distance. More specifically, from now on let the laws $\mathbb{P},\mathbb{Q} \in \mathcal{M}_1(\mathbb{R}^d)$ be in the class of d-variate normals, i.e.

$$\mathbb{P} \sim \mathcal{N}(\mu,\Sigma) \quad \text{and} \quad \mathbb{Q} \sim \mathcal{N}(\nu,\Xi) \quad \text{for some } \mu,\nu \in \mathbb{R}^d,\ \Sigma,\Xi \in S_+(\mathbb{R}^d), \tag{4}$$

the symmetric, positive definite, d-dimensional matrices. From now on we will also restrict to $p = 2$. In this case the Wasserstein distance between $\mathcal{N}(\mu,\Sigma)$ and $\mathcal{N}(\nu,\Xi)$ is computed as (see [19,35,27])

$$\mathcal{GW} := \mathcal{W}_2^2(\mathbb{P},\mathbb{Q}) = \|\mu-\nu\|^2 + \operatorname{tr}(\Sigma) + \operatorname{tr}(\Xi) - 2\operatorname{tr}\left[\left[\Sigma^{1/2}\,\Xi\,\Sigma^{1/2}\right]^{1/2}\right]. \tag{5}$$

Here, tr refers to the trace of a matrix and its square root is defined in the usual spectral way. The norm $\|\cdot\|$ is the Euclidean norm with corresponding scalar product denoted by $\langle\cdot,\cdot\rangle$. Now, if we replace $\mathbb{P}$ with the empirical measure $\mathbb{P}_n$ and read μ and Σ as a functional of $\mathbb{P}$, we obtain the empirical Wasserstein estimator $\widehat{\mathcal{GW}}_n$ restricted to the d-dimensional Gaussian measures as

$$\begin{aligned} \widehat{\mathcal{GW}}_n &= \widehat{\mathcal{GW}}_n(X_1,\dots,X_n,\mathbb{Q}) \\ &:= \mathcal{W}_2^2\left(\mathcal{N}(\hat{\mu}_n,\hat{\Sigma}_n),\mathcal{N}(\nu,\Xi)\right) \\ &= \|\hat{\mu}-\nu\|^2 + \operatorname{tr}(\hat{\Sigma}) + \operatorname{tr}(\Xi) - 2\operatorname{tr}\left[\left[\hat{\Sigma}^{1/2}\,\Xi\,\hat{\Sigma}^{1/2}\right]^{1/2}\right]. \end{aligned} \tag{6}$$

Similar to the case of the general empirical Wasserstein distance for $d = 1$ we find in the following that the asymptotic behavior differs whether $\mathbb{P} = \mathbb{Q}$, i.e., $\mu = \nu$ and $\Sigma = \Xi$ or $\mathbb{P} \neq \mathbb{Q}$. Let us start with the latter case which turns out to be

Download English Version:

https://daneshyari.com/en/article/1145174

Download Persian Version:

https://daneshyari.com/article/1145174

Daneshyari.com