WORKING PAPERS

N° 1039

September 2019

"Functional, randomized and smoothed multivariate quantile regions"

Olivier P. Faugeras and Ludger Rüschendorf

Toulouse School of Economics

Functional, randomized and smoothed multivariate quantile regions

Olivier P. Faugeras
Toulouse School of Economics,
University of Toulouse Capitole

Ludger Rüschendorf,
Abteilung für Mathematische Stochastik,
Albert-Ludwigs University of Freiburg

July 11, 2019

Abstract

A notion of multivariate depth, resp. quantile region, was introduced in [Chernozhukov et al., 2017], based on a mass transportation approach. In [Faugeras and Rüschendorf, 2017], this approach was generalized by defining quantiles as Markov morphisms carrying suitable algebraic, ordering and topological structures over probability measures. In addition, a copula step was added to the mass transportation step. Empirical versions of these depth areas do not give exact level depth regions. In this paper, we introduce randomized depth regions by means of a formulation by depth functions, resp. by randomized quantiles sets. These versions attain the exact level and also provide the corresponding consistency property. We also investigate in the case of continuous marginals a smoothed version of the empirical copula and compare its behavior with the unsmoothed version. Extensive simulations illustrate the resulting randomized depth areas and show that they give a valid representation of the central depth areas of a multivariate distribution, and thus are a valuable tool for their analysis.

1 Introduction

1.1 The combined copula-mass transportation approach to multivariate quantiles and depth areas

In [Chernozhukov et al., 2017], a mass transportation approach has been proposed to the definition of multivariate quantiles and depth areas. The basic idea in their paper is that balls give a natural definition of central regions of a spherical distribution. By mass transportation, these central ball regions are

mapped to central center-outward quantile domains, i.e. depth regions, in the observation domain.

Mass transportation can be regarded as a quantitative approach for transforming measures. Under regularity assumptions, the optimal mass transportation is indeed induced by mappings– in general, it is induced by Markov kernels. Transformation of measures by mappings can also be studied from a more qualitative and geometric viewpoint, by pushing forward a non-atomic measure by a cyclically monotone mapping, see [Smith and Knott, 1987], [Rüschendorf and Rachev, 1990], [McCann, 1995]. This was used in [Hallin, 2017] and [del Barrio et al., 2018] to defining a multivariate quantile function only at the observed sample points by solving an empirical optimal matching problem. Note that the idea of defining a multivariate quantile by a push-forward from a reference distribution can be traced back to early ideas in [Easton and McCulloch, 1990], where the authors looked for the optimal matching between a sample of observed values and a dataset sampled from a reference distribution, in order to construct multivariate $Q-Q$ plots.

In [Faugeras and Rüschendorf, 2017], the mass transportation approach of [Chernozhukov et al., 2017] and [del Barrio et al., 2018] was generalized by defining a quantile as a Markov kernel between such a reference spherical distribution and the multivariate distribution under consideration, compatible with corresponding algebraic, ordering and topological structures. In addition, a copula step is introduced, so that the regularity assumptions are satisfied and that the optimal transportation step, now between the reference spherical measure and the copula measure, is indeed induced by a mapping.

More precisely, the setting and notations of [Chernozhukov et al., 2017], [del Barrio et al., 2018], and [Faugeras and Rüschendorf, 2017] are as follows: for a multivariate random vector $\mathbf{X} \in \mathbb{R}^d$ with c.d.f. F, we denote by $\mathbf{G} = (G_1, \ldots, G_d)$ the vector of marginal cdfs, i.e. $G_i(x_i) := F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$. We consider the spherical distribution $P^{\mathbf{S}}$ of a r.v. $\mathbf{S}$ on the unit ball $\mathbf{B}_1 = \{\mathbf{s} \in \mathbb{R}^d; ||\mathbf{s}|| \leq 1\}$. Identifying $\mathbf{s} \in \mathbf{B}_1$ with the pair $(r, \mathbf{a})$, where the radial part $r \sim U_{[0,1]}$ is uniformly distributed on the interval $[0,1]$, and the angular part $\mathbf{a}$ is uniformly distributed on the unit sphere in $\mathbb{R}^d$, such a spherical distribution $P^{\mathbf{S}}$ gives natural depth regions of $P^{\mathbf{S}}$ mass τ as the balls $\mathbf{B}_\tau$ of radius τ, since $P^{\mathbf{S}}(\mathbf{B}_\tau) = \tau$. The basic idea of [Chernozhukov et al., 2017] and [del Barrio et al., 2018] is then to transform these balls B_τ into depth regions of level τ in the observational space, by setting

$$\mathbf{A}_\tau := \mathbf{Q}_F(\mathbf{B}_\tau), \tag{1}$$

where $\mathbf{Q}_F$ is the multivariate "quantile" optimal mass transportation mapping obtained by pushing forward $P^{\mathbf{S}}$ to $P^{\mathbf{X}}$ by optimal transport, i.e.

$$\mathbf{Q}_F \# P^{\mathbf{S}} = P^{\mathbf{X}}. \tag{2}$$

$\mathbf{Q}_F$ is the optimal mapping in the solution of the Monge-Kantorovich Wasserstein metric,

$$W_2^2(P^{\mathbf{S}}, P^{\mathbf{X}}) = \inf \int ||\mathbf{x} - \mathbf{y}||^2 \gamma(\mathbf{dx}, \mathbf{dy}),$$

where the infimum is over all couplings γ of $(P^{\mathbf{S}}, P^{\mathbf{X}})$. By noting $\mathbf{R}_F$ the corresponding "rank" reciprocal optimal mass transportation mapping, i.e.

$$\mathbf{R}_F \# P^{\mathbf{X}} = P^{\mathbf{S}}, \tag{3}$$

one gets that

$$\begin{aligned} P^{\mathbf{X}}(\mathbf{A}_\tau) &= P(\mathbf{Q}_F(\mathbf{S}) \in \mathbf{Q}_F(\mathbf{B}_\tau)) \\ &= P(\mathbf{S} \in \mathbf{R}_F \circ \mathbf{Q}_F(\mathbf{B}_\tau)) = P^{\mathbf{S}}(\mathbf{B}_\tau) = \tau. \end{aligned} \tag{4}$$

For (2), (3) and thus (4) to hold, one needs moment and regularity conditions on the distribution $P^{\mathbf{X}}$ (see [Rüschendorf and Rachev, 1990], [Brenier, 1991], and [McCann, 1995], where the moment assumption is dropped). These regularity assumptions are e.g. not satisfied when $P^{\mathbf{X}}$ has a discrete component, so, in particular, when $P^{\mathbf{X}}$ is replaced by the empirical measure $P^{\mathbf{X}_n^*} := n^{-1} \sum_{i=1}^n \delta_{\mathbf{X}_i}$ based on an ergodic sample realization $\mathbf{X}_1, \mathbf{X}_2, \ldots$ of $\mathbf{X}$. Thus, the empirical versions of the depth areas in [Chernozhukov et al., 2017] require a smoothed version of the empirical measure and regularity assumptions. In order to alleviate these issues, and also for obtaining depth areas fully equivariant w.r.t. to marginal monotone transformations, [Faugeras and Rüschendorf, 2017] introduce a preliminary copula step based on the distributional transform and do instead the optimal transportation step between the reference spherical measure $P^{\mathbf{S}}$ and the copula measure. This combined copula-mass transportation approach modifies (1) and is summarized as follows:

1. Transform $\mathbf{X} \sim F$ into its copula representer $\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$, whose c.d.f. C is the copula of F. Here, $\mathbf{V}$ is a vector with i.i.d. components uniformly distributed on $[0, 1]$, and $\mathbf{X} \to \mathbf{G}(\mathbf{X}, \mathbf{V})$ is the multivariate marginal distributional transform, see [Faugeras and Rüschendorf, 2017]. Note that, under the assumption that $P^{\mathbf{X}}$ is either discrete or absolutely continuous, $P^{\mathbf{U}}$ is always absolutely continuous, see Lemma 6.1 in Appendix 6.1. Hence, the regularity assumptions are satisfied for that the optimal transportation plans between $P^{\mathbf{U}}$ and $P^{\mathbf{S}}$ be induced by mappings.

2. Transport $P^{\mathbf{U}}$ into the spherical reference distribution $P^{\mathbf{S}}$ via optimal mass transportation maps $\mathbf{Q}_C, \mathbf{R}_C$, i.e.

$$\mathbf{Q}_C \# P^{\mathbf{S}} = P^{\mathbf{U}}, \quad \mathbf{R}_C \# P^{\mathbf{U}} = P^{\mathbf{S}}. \tag{5}$$

3. The balls $\mathbf{B}_\tau$ of $P^{\mathbf{S}}$-mass τ, are mapped into depths regions $\mathbf{A}_\tau$ at the copula level of $P^{\mathbf{U}}$-mass τ:

$$\mathbf{A}_\tau := \mathbf{Q}_C(\mathbf{B}_\tau), \quad P^{\mathbf{U}}(\mathbf{A}_\tau) = \tau. \tag{6}$$

4. Use the multivariate marginal quantile transform $\mathbf{G}^{-1} = (G_1^{-1}, \ldots, G_d^{-1})$, whose components are the inverse marginal distribution functions, to define depth regions $\mathbf{Z}_\tau$ in the observational space as

$$\mathbf{Z}_\tau := \mathbf{G}^{-1}(\mathbf{A}_\tau). \tag{7}$$

Under regularity assumptions on the marginal c.d.f.s. (e.g. each marginal c.d.f. is continuous and strictly increasing), one obtains depth regions of level τ at the observational $P^{\mathbf{X}}$ level, i.e.

$$P^{\mathbf{X}}(\mathbf{Z}_\tau) = \tau.$$

The procedure can be formulated in terms of transformations of probability measures via Markov kernels, as in Figure 1.

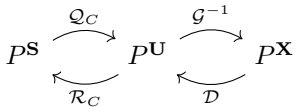

Figure 1: Markov morphisms of probability measures

We denote by $\mathcal{Q}_C, \mathcal{R}_C, \mathcal{G}^{-1}$ the degenerate Markov morphisms induced by the mappings $\mathbf{Q}_C, \mathbf{R}_C, \mathbf{G}^{-1}$, e.g. $\mathcal{Q}_C(\mathbf{s}, .) = \delta_{\mathbf{Q}_C(\mathbf{s})}(.)$. $\mathcal{D}$ denotes the (non-degenerate) Markov kernel corresponding to the distributional transform $\mathbf{X} \to \mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$, i.e. $D(\mathbf{x}, .) = P^{\mathbf{U}|\mathbf{X}=\mathbf{x}}(.) = P^{\mathbf{G}(\mathbf{x},\mathbf{V})}(.)$, since $\mathbf{V}$ is chosen independent of $\mathbf{X}$.

1.2 Inspection of empirical depth areas

The empirical version of these central quantile areas requires a modified treatment, as we now proceed. It is described by the following diagram in Figure 2, with notations similar as in Figure 1:

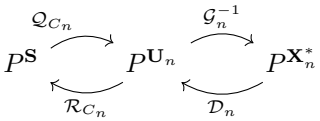

Figure 2: Markov morphisms of empirical probability measures

Here, we have:

1. $\mathbf{X}_n^* \sim F_n$ is, conditionally on the sample, a bootstrap replication, distributed according to the empirical c.d.f. F_n. $\mathbf{U}_n = \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$ is the empirical copula representer, whose c.d.f. is the empirical copula function C_n, obtained by the empirical distributional transform with corresponding Markov kernel $\mathcal{D}_n$.

2. $\mathbf{Q}_{C_n}, \mathbf{R}_{C_n}$ are the optimal transportation maps between the empirical copula and reference measures, viz.

$$\mathbf{Q}_{C_n} \# P^{\mathbf{S}} = P^{\mathbf{U}_n}, \quad \mathbf{R}_{C_n} \# P^{\mathbf{U}_n} = P^{\mathbf{S}}. \tag{8}$$

3. Defining, similarly to (6) and (7) the depth regions

$$\mathbf{A}_n := \mathbf{Q}_{C_n}(\mathbf{B}_\tau), \quad \mathbf{Z}_n := \mathbf{G}_n^{-1}(\mathbf{A}_n), \tag{9}$$

for a fixed level $0 < \tau < 1$, one obtains $P^{\mathbf{U}_n}(\mathbf{A}_n) = \tau$, but in general $\mathbf{Z}_n$ is not of exact $P^{\mathbf{X}_n^*}$-level τ. If τ is not in the range of values $\{0, 1/n, \ldots, 1\}$ of the empirical measure $P^{\mathbf{X}_n^*}$, it is clear that $\mathbf{Z}_n$ can not be a set of $P^{\mathbf{X}_n^*}$-mass τ. A more fundamental reason comes from the fact that since $\mathcal{D}$ (in the case of a discrete $\mathbf{X}$) and $\mathcal{D}_n$ (in every case) are non-degenerate Markov kernels (due to the presence of the randomizer $\mathbf{Z}$ in the distributional transforms), sets like $\mathbf{A}$ (resp. $\mathbf{A}_n$) are no longer transformed into sets by the contravariant action of the Markov kernels D (resp. $\mathcal{D}_n$), but into randomized sets or functions, see Sections 2 and 3 for details.

1.3 Outline

As a consequence of the above given discussion, for a general distribution and also at the empirical level, the natural constructions (9) lead to depth domains of inexact level, which also induces problems in the proof of the consistency result of Corollary 6.4 in [Faugeras and Rüschendorf, 2017]. In this paper, we show how these issues can be circumvented by three possible approaches: in the first approach in Section 2, we abandon the idea of having depth areas of given exact level τ in the observation space, getting only depth areas at the copula level, and introducing instead *membership functions*. These membership functions generalize indicator functions of sets and specify for any point $\mathbf{x}$ with which probability it is contained in a corresponding *randomized* depth area. This allows to turn depth area at the level of the reference distribution $P^{\mathbf{S}}$ into *depth region functions* at the observation level. In Section 3, we introduce *randomized depth areas* as random sets, which attain the exact level τ. Both descriptions are closely connected and allow to deal with the empirical versions as well as to prove consistency results. In Section 4, we circumvent the issues related to the nondegeneracy of the distributional transform kernel by introducing a preliminary smoothing of the empirical measure. For a continuous $P^{\mathbf{X}}$, this allows to retain the advantage of the copula + mass transportation approach while obtaining only degenerate Markov kernels and depth areas as proper deterministic sets. Eventually, we show in Section 5 by simulations how the different versions of the depth areas introduced can be used as a data analytical tool for the analysis of multivariate distributions. An auxiliary result of independent interest on the absolute continuity of copula measures of absolutely continuous distributions is proven in Appendix 6.1. Proofs of the main results are relegated to Appendix 6.2 and technical details of the simulations and supplementary numerical illustrations are to be found in Appendix 7.

1.4 Setting and notation

We write $\mathbb{R}^d$-valued elements, like random vectors $\mathbf{X}$ or sets $\mathbf{A}$, in bold letters, and interpret operations between vectors componentwise. $P^{\mathbf{X}}$ will stand for the

law associated with its representing variable $\mathbf{X}$. We follow the framework and assumptions established in [Faugeras and Rüschendorf, 2017]:

- **Ergodicity hypothesis:**
  let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be an ergodic sample realization of $P^{\mathbf{X}}$ defined on some probability space $(\Omega, \mathcal{A}, P)$. It will be understood that all random variables defined in this article, $\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{V}$, and $\mathbf{S}_n, \mathbf{U}_n, \mathbf{X}_n^*$ of Sections 2-4, $\hat{\mathbf{S}}_n, \hat{\mathbf{U}}_n, \hat{\mathbf{X}}_n^*$ of Section 4, are defined on the auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$ on which Skorohod's Theorem in Theorem 6.2 in [Faugeras and Rüschendorf, 2017] holds, i.e. on the auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$ which allows to construct representers $\mathbf{X}_n^*$ of the empirical measure, $\mathbf{X}$ of $P^{\mathbf{X}}$ s.t., with P-probability one has

  $$\mathbf{X}_n^* \overset{P^*a.s.}{\longrightarrow} \mathbf{X}, \tag{10}$$

  see steps one and two in the proof of Theorem 6.2 in [Faugeras and Rüschendorf, 2017]. (Compared to the notation in [Faugeras and Rüschendorf, 2017], we simplify notation and drop the * in the $\mathbf{X}^*$, which reminded that $\mathbf{X}^*$ representing $P^{\mathbf{X}}$ was defined on this auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$). We will denote, for simplicity, e.g. by $P^{\mathbf{S}}$ the law of $\mathbf{S}$, and not by $P^{*\mathbf{S}}$.

- **Regularity assumption on $P^{\mathbf{X}}$:**
  Unless stated otherwise, $P^{\mathbf{X}}$ will be assumed either discrete or absolutely continuous w.r.t. λ^d, the $d-$dimensional Lebesgue measure.

2 Depth areas defined by membership functions

2.1 Membership functions

At the categorical level, the issue mentioned in the introduction arises from the fact that the morphisms of the category of measurable spaces are the measurable mappings, whereas in the category of (probability) measures (or dually in the category of measurable functions), the morphisms are the Markov kernels. Only kernels $\mathcal{K} : P^{\mathbf{X}} \to P^{\mathbf{Y}} := \mathcal{K}P^{\mathbf{X}}$, arising from a transformation of a non-atomic probability measure $P^{\mathbf{X}}$, can be identified with a measurable mapping $f : (\mathcal{X}, \mathcal{A}(\mathcal{X})) \to (\mathcal{Y}, \mathcal{A}(\mathcal{Y}))$ between the underlying measurable spaces, as $\mathcal{K}$ becomes in this case degenerate, i.e.

$$\mathcal{K}(\mathbf{x}, d\mathbf{y}) = \delta_{f(\mathbf{x})}(\mathbf{dy}). \tag{11}$$

The consequence for sets and their transformations is the following: whereas measurable sets are transformed into measurable sets by the morphisms– measurable mappings, in the category of measurable spaces, in the category of measurable functions (dual to the category of probability measures), non-degenerate Markov kernels $\mathcal{K}$ transforms measurable sets $\mathbf{A} \in \mathcal{A}(Y)$ into measurable *non-binary functions* $\mathbf{x} \to \mathcal{K}(\mathbf{x}, \mathbf{A})$. Indeed, sets $\mathbf{A} \in \mathcal{A}(Y)$ can be identified with

their indicator functions $\mathbf{y} \to \mathbb{1}_{\mathbf{A}}(\mathbf{y})$ whose contravariant transformation by the Markov kernel $\mathcal{K}$ yields the function

$$\mathbf{x} \to \mathcal{K}(\mathbf{x}, \mathbf{A}) = \int_{\mathbf{A}} \mathcal{K}(\mathbf{x}, \mathbf{dy}). \tag{12}$$

When $\mathcal{K}$ is degenerate of the form (11), (12) writes

$$\mathcal{K}(\mathbf{x}, \mathbf{A}) = \int_{\mathbf{A}} \delta_{f(\mathbf{x})}(\mathbf{dy}) = \mathbb{1}_{f^{-1}(\mathbf{A})}(\mathbf{x})$$

and the latter, as a binary function, can be identified with the measurable set $f^{-1}(\mathbf{A}) \in \mathcal{A}(\mathcal{X})$. For a non-degenerate Markov kernel, this is no longer the case, as the function (12) is in general non-binary. Therefore, one can not hope to obtain depth sets from transformations by non-degenerate kernels.

These considerations suggest that a first approach to introduce the equivalent of depth areas of exact level τ is to consider depth functions instead of depth sets. This is done by embedding sets into indicator functions, and considering transformation by Markov kernels of these indicator functions. To that purpose, we introduce membership functions which describe for a point $\mathbf{x}$ in the space considered the probability with which $\mathbf{x}$ belongs to a (randomized) depth areas.

The precise formulation of this idea is as follows. Let us fix the level $0 < \tau < 1$ throughout the paper. We will simplify notations and simply write the sets $\mathbf{B}_\tau, \mathbf{A}_\tau, \mathbf{Z}_\tau$ of (6), (7) without the subscript τ. For the set $\mathbf{B} := \mathbf{B}_\tau$ in the spherical reference space $\mathbf{B}_1$, with $P^{\mathbf{S}}(\mathbf{B}) = \tau$, we define the corresponding membership function b by the indicator function,

$$b(\mathbf{s}) = \mathbb{1}_{\mathbf{B}}(\mathbf{s}), \quad \mathbf{s} \in \mathbf{B}_1.$$

At the level of $\mathbf{S}$, $\mathbf{B}$ is a set of level τ, viz.

$$P^{\mathbf{S}}(b) = \int b(\mathbf{s}) P^{\mathbf{S}}(\mathbf{ds}) = P(\mathbf{S} \in \mathbf{B}) = \tau.$$

We get the membership function a on the copula space $[0,1]^d$ and z on the sample space $\mathbb{R}^d$ by letting the Markov kernels act in a contravariant fashion:

$$a := \mathcal{R}_C(b), \quad z := \mathcal{D}a = \mathcal{D}\mathcal{R}_C b, \tag{13}$$

i.e.

$$a(\mathbf{u}) = \int \mathcal{R}_c(\mathbf{u}, \mathbf{ds}) b(\mathbf{s}), \tag{14}$$

and

$$z(\mathbf{x}) = \int \mathcal{D}(\mathbf{x}, \mathbf{du}) a(\mathbf{u}) = \int \left(\int \mathcal{R}_c(\mathbf{u}, \mathbf{ds}) b(\mathbf{s}) \right) \mathcal{D}(\mathbf{x}, \mathbf{du}). \tag{15}$$

In diagram form, these relations are represented as in Figure 3, the dual diagram of Figure 1.

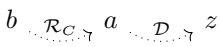

Figure 3: Depth regions functions–population version

For the empirical versions of the depth areas in terms of membership functions, one defines similarly,

$$a_n := \mathcal{R}_{C_n} b, \quad z_n := \mathcal{D}_n a_n = \mathcal{D}_n \mathcal{R}_{C_n} b, \tag{16}$$

as shown in Figure 4.

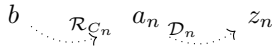

Figure 4: Depth regions functions–empirical version

Since $\mathcal{R}_C$, $\mathcal{R}_{C_n}$ are degenerate Markov kernels induced by the mappings $\mathbf{R}_C$, $\mathbf{R}_{C_n}$, with reciprocal functions $\mathbf{Q}_C$, $\mathbf{Q}_{C_n}$, the membership functions a, a_n are proper indicator functions defining proper subsets in the copula space

$$a(\mathbf{u}) = \mathbb{1}_{\mathbf{Q}_C(\mathbf{B})}(\mathbf{u}) = \mathbb{1}_{\mathbf{A}}(\mathbf{u}), \quad a_n(\mathbf{u}) = \mathbb{1}_{\mathbf{Q}_{C_n}(\mathbf{B})}(\mathbf{u}) = \mathbb{1}_{\mathbf{A}_n}(\mathbf{u}), \tag{17}$$

with $\mathbf{A}, \mathbf{A}_n$ previously defined as in (6) and (9). Indeed, by (13),

$$\begin{aligned} a(\mathbf{u}) &= \int \mathcal{R}_C(\mathbf{u}, \mathbf{ds}) \mathbb{1}_{\mathbf{B}}(\mathbf{s}) = \int_{\mathbf{B}} \delta_{\mathbf{R}_C(\mathbf{u})}(\mathbf{ds}) \\ &= \mathbb{1}_{\mathbf{R}_C(\mathbf{u}) \in \mathbf{B}} = \mathbb{1}_{\mathbf{u} \in \mathbf{R}_C^{-1}(\mathbf{B})} = \mathbb{1}_{\mathbf{u} \in \mathbf{Q}_C(\mathbf{B})} = \mathbb{1}_{\mathbf{A}}(\mathbf{u}) \end{aligned}$$

Thus the membership functions a, a_n can be identified with the sets $\mathbf{A} = \mathbf{Q}_C(\mathbf{B})$, $\mathbf{A}_n = \mathbf{Q}_{C_n}(\mathbf{B})$, images of the depth ball $\mathbf{B}$ under the transportation maps $\mathbf{Q}_C, \mathbf{Q}_{C_n}$.

In contrast to this case, in the case of non-Lebesgue continuous $P^{\mathbf{X}}$, $\mathcal{D}$ is a non-degenerate Markov kernel and, in general, the empirical $\mathcal{D}_n$ is always non-degenerate. Therefore, the membership functions z, z_n of (13) and (16) are no longer indicator functions of some deterministic depth set, but only determines a "random" depth sets by their membership probabilities, as is explained below.

2.2 Interpretation of the membership functions z, z_n

As explained in Remark 4.5 in [Faugeras and Rüschendorf, 2017], on Polish spaces , non-degenerate Markov morphisms such as $\mathcal{D}, \mathcal{D}_n$ can be considered as degenerate Markov morphisms on an enlarged probability space. In the case of $\mathcal{D}, \mathcal{D}_n$ this representation is implicit in the construction of $\mathcal{D}, \mathcal{D}_n$ as random transformations $\mathbf{X} \to \mathbf{G}(\mathbf{X}, \mathbf{V})$ and $\mathbf{X}_n^* \to \mathbf{G}_n(\mathbf{X}_{\mathbf{n}}^*, \mathbf{V})$, with independent randomizer $\mathbf{V} \in [0,1]^d$. In other words, if one considers the enlarged joint

distributions $P^{(\mathbf{X},\mathbf{V})} = P^{\mathbf{X}} \otimes P^{\mathbf{V}}$, $P^{(\mathbf{U},\mathbf{V})}$ (and similarly $P^{(\mathbf{X}_\mathbf{n}^*,\mathbf{V})}$, $P^{(\mathbf{U}_\mathbf{n},\mathbf{V})}$ for the empirical counterpart), the corresponding Markov morphisms transforming these probability measures are degenerate and induced by the pair of joint mappings

$$\begin{array}{ccc} (\mathbf{G}(\mathbf{X},\mathbf{V}),\mathbf{V}) & \overset{(\mathbf{G},\mathbf{id})}{\longleftarrow} & (\mathbf{X},\mathbf{V}) \\ (\mathbf{U},\mathbf{V}) & \overset{(\mathbf{G}^{-1},\mathbf{id})}{\longrightarrow} & (\mathbf{X},\mathbf{V}). \end{array}$$

Dually, this means that the membership functions z, z_n obtained from $\mathcal{D}, \mathcal{D}_n$ can be construed as indicator functions of sets on the enlarged $(\mathbf{X},\mathbf{V})$ space $\mathbb{R}^d \times [0,1]^d$. More precisely, one has that for the multivariate distributional transform with $\mathbf{V}$ made of independent components, $\mathcal{D}(\mathbf{x},.)$ is the uniform distribution on the box $[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})]$, if $\mathbf{G}(\mathbf{x}-) \neq \mathbf{G}(\mathbf{x})$:

$$\mathcal{D}(\mathbf{x},\mathbf{du})) = \frac{\mathbf{du}}{m(\mathbf{x},\mathbf{G})} \mathbb{1}_{[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})]}(\mathbf{u}), \quad \text{for } \mathbf{G}(\mathbf{x}) \neq \mathbf{G}(\mathbf{x}-),$$

where we have set $m(\mathbf{x},\mathbf{G}) := \Pi_{i=1}^d G_i(x_i) - G_i(x_i-)$. As a consequence, the image of an indicator function $\mathbb{1}_{[\mathbf{u_1},\mathbf{u_2}]}$ by $\mathcal{D}$ becomes, for $\mathbf{G}(\mathbf{x}) \neq \mathbf{G}(\mathbf{x}-)$ and $[\mathbf{u_1},\mathbf{u_2}] \subset [\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})]$,

$$\begin{aligned} (\mathcal{D}\mathbb{1}_{[\mathbf{u_1},\mathbf{u_2}]})(\mathbf{x}) &= \int \frac{\mathbf{du}}{m(\mathbf{x},\mathbf{G})} \mathbb{1}_{[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})]}(\mathbf{u}) \mathbb{1}_{[\mathbf{u_1},\mathbf{u_2}]}(\mathbf{u}) \\ &= \frac{\lambda^d([\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})] \cap [\mathbf{u_1},\mathbf{u_2}])}{m(\mathbf{x},\mathbf{G})}, \quad \text{for } \mathbf{G}(\mathbf{x}) \neq \mathbf{G}(\mathbf{x}-), \end{aligned}$$

where λ^d is the $d-$variate Lebesgue measure on $[0,1]^d$ (i.e. the distribution of $\mathbf{V}$), see Figure 5.

Thus, the z function, which is the image of $\mathbb{1}_\mathbf{A}$ by $\mathcal{D}$, writes at a jump point $\mathbf{x}$ with positive mass as

$$z(\mathbf{x}) = \frac{\lambda^d([\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})] \cap \mathbf{A})}{m(\mathbf{x},\mathbf{G})}, \quad \text{for } \mathbf{G}(\mathbf{x}) \neq \mathbf{G}(\mathbf{x}-),$$

and similarly for z_n. As a conclusion, the functions z, z_n have an interpretation as a sort of generalized indicator function: it is one if $[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})] \subset \mathbf{A}$, zero if $[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})] \cap \mathbf{A} = \varnothing$ and amounts to the relative portion of mass of $\mathbf{A}$ in $[\mathbf{G}(\mathbf{x}-),\mathbf{G}(\mathbf{x})]$.

In particular, for the empirical membership function z_n, one can construe it as a subset of the original cloud of sample points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, each attached with a portion of the "ghost" $[0,1]^d$ space, standing for the amount of mass required to obtain $P^{\mathbf{X}_n^*}$ expectation τ: it can be phenomenologically interpreted as sort of a random set of $\mathbb{R}^d$, where each point is selected with a randomization weight corresponding to a fraction of the volume in the enlargement space $[0,1]^d$. The precise description of these enlarged sets of $\mathbb{R}^d \times [0,1]^d$ as random sets of $\mathbb{R}^d$ will be given in Section 3.

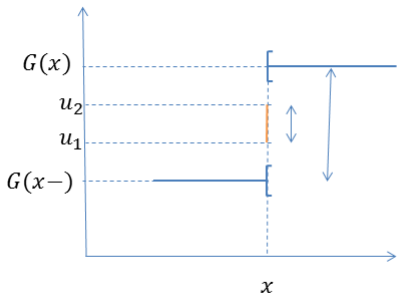

Figure 5: Interpretation of the membership functions z. We represent in the univariate case, i.e. for one marginal $G := G_i$ of $\mathbf{G}$, the transformation of an indicator function $\mathbb{1}_{[u_1,u_2]}$ in the copula space at a discontinuity point x of $\mathbf{G}$ by the Markov morphism $\mathcal{D}$. One obtains the fraction of mass in orange of the interval $[u_1, u_2]$, divided by the marginal mass of the jump $G(x) - G(x-)$. In the multidimensional case, one obtains a similar, yet harder to visualize, figure, where $z(\mathbf{x})$ stands for the d-dimensional Lebesgue measure of the box $[\mathbf{u_1}, \mathbf{u_2}]$ divided by the product of marginals jump mass $m(\mathbf{x}, \mathbf{G})$.

2.3 Properties of depth region and membership functions

The following theorem clarifies what can be expected for the depth regions and membership functions, and corrects Corollary 6.4 in [Faugeras and Rüschendorf, 2017]:

Theorem 2.1. *With $P-$probability one,*

1. *At the copula level:*

   (a) $\mathbf{A}$ *is exactly of* $P^{\mathbf{U}}$*-mass* τ, $\mathbf{A}_n$ *is exactly of* $P^{\mathbf{U}_n}$*-mass* τ*:*

   $$P^{\mathbf{U}}(\mathbf{A}) = P^{\mathbf{U}_n}(\mathbf{A}_n) = \tau.$$

   (b) *With $P-$probability one, the empirical depth area $\mathbf{A}_n$ at the copula level is asymptotically of $P^{\mathbf{U}}$-mass τ:*

   $$P^{\mathbf{U}}(\mathbf{A}_n) \to \tau, \quad \text{as } n \to \infty.$$

   (c) *With $P-$probability one, the $L_1(P^{\mathbf{U}})$ distance between the membership functions a_n and a, (equivalently the $P^{\mathbf{U}}$ symmetric distance between the depth sets $\mathbf{A}_n$ and $\mathbf{A}$) is asymptotically null:*

   $$P^{\mathbf{U}}(|a_n - a|) = P^{\mathbf{U}}(\mathbf{A}_n \Delta \mathbf{A}) \to 0, \tag{18}$$

   as $n \to \infty$, with $P-$probability one.

2. *At the observational* $\mathbf{X}$ *level,*

   (a) *the depth set* $\mathbf{Z}$ *is of* $P^{\mathbf{X}}$*-mass at least* τ*, the depth set* $\mathbf{Z}_n$ *is of* $P^{\mathbf{X}_n^*}$*-mass at least* τ*,*

$$P^{\mathbf{X}}(\mathbf{Z}) \geq \tau, \quad P^{\mathbf{X}_n^*}(\mathbf{Z}_n) \geq \tau.$$

   while the membership functions z, z_n *are of exactly* $P^{\mathbf{X}}, P^{\mathbf{X}_n^*}$*-expectation* τ*:*

$$P^{\mathbf{X}}(z) = \tau, \quad P^{\mathbf{X}_n^*}(z_n) = \tau.$$

   If $P^{\mathbf{X}}$ *is continuous, then* $\mathbf{Z}$ *is exactly of* $P^{\mathbf{X}}$*-mass* τ*:* $P^{\mathbf{X}}(\mathbf{Z}) = \tau$*.*

   (b) *With* $P-$*probability one, if* $P^{\mathbf{X}}$ *is continuous, the* $P^{\mathbf{X}}$ *symmetric distance between* $\mathbf{Z}$ *and* $\mathbf{Z}_n$ *becomes asymptotically negligible,*

$$P^{\mathbf{X}}(\mathbf{Z}_n \Delta \mathbf{Z}) \to 0,$$

   so that $\mathbf{Z}_n$ *is asymptotically of* $P^{\mathbf{X}}$*-mass* τ*:*

$$P^{\mathbf{X}}(\mathbf{Z}_n) \to \tau.$$

   (c) *With* $P-$*probability one, the* $L_1(P^{\mathbf{X}})$ *distance between the membership functions* z_n *and* z*, is asymptotically null:*

$$P^{\mathbf{X}}(|z - z_n|) \to 0,$$

   as $n \to \infty$*, with* P*-probability one.*

3 Randomized depths areas

The discussion in Section 2 of the Markov kernels $\mathcal{D}$, resp. $\mathcal{D}_n$, corresponding to the multivariate distributional transforms $\mathbf{G}(\mathbf{X}, \mathbf{V})$, resp. $\mathbf{G_n}(\mathbf{X_n^*}, \mathbf{V})$, gives us a tool to define directly a random depth area as a random set.

3.1 Definitions

Let the level $0 < \tau < 1$ remain fixed as before. Define, for $\mathbf{v} \in [0,1]^d$, the set

$$\mathbf{Z_v} := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{G}(\mathbf{x}, \mathbf{v}) \in \mathbf{A}\}. \tag{19}$$

The randomized depth area at level τ is then defined as the random set $\mathbf{Z_V}$, where $\mathbf{V}$ is the randomizer used in the distributional transforms $\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$ and $\mathbf{U}_n = \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$.

These randomized depth regions are random subsets of $\mathbb{R}^d$ of exact mass τ: By definition, we get

$$\begin{aligned} P^{\mathbf{X}}(\mathbf{Z_V}) &= P^*(\mathbf{X} \in \mathbf{Z_V}) \\ &= P^{(\mathbf{X},\mathbf{V})}(\{(\mathbf{x}, \mathbf{v}) : \mathbf{x} \in \mathbf{Z_v}\}) \\ &= P^{(\mathbf{X},\mathbf{V})}(\{(\mathbf{x}, \mathbf{v}) : \mathbf{G}(\mathbf{x}, \mathbf{v}) \in \mathbf{A}\}) \\ &= P^*(\mathbf{G}(\mathbf{X}, \mathbf{V}) \in \mathbf{A}) = P^*(\mathbf{U} \in \mathbf{A}) \\ &= P^*(\mathbf{S} \in \mathbf{B}) = \tau. \end{aligned}$$

Similarly, for their empirical counterparts, set

$$\mathbf{Z}_{\mathbf{v},n} := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{G_n}(\mathbf{x}, \mathbf{v}) \in \mathbf{A}_n\}. \tag{20}$$

and define the empirical randomized depth region as $\mathbf{Z}_{\mathbf{V},n}$. These empirical randomized depth regions are also random subsets of $\mathbb{R}^d$ of exact mass τ:

$$P^*(\mathbf{X}_n^* \in \mathbf{Z}_{\mathbf{V},n}) \quad = \quad P^*(\mathbf{G_n}(\mathbf{X_n^*}, \mathbf{V}) \in \mathbf{A}_n) = P^*(\mathbf{U}_n \in A_n) = \tau.$$

3.2 Convergence of the randomized depth region

The consistency properties of these empirical randomized depth areas is stated in the following theorem:

Theorem 3.1. *The empirical randomized depth area $\mathbf{Z}_{\mathbf{V},n}$ is asymptotically consistent in the $P^{\mathbf{X}}$-symmetric difference distance towards $\mathbf{Z}_{\mathbf{V}}$: one has*

$$P^{\mathbf{X}}(\mathbf{Z}_{\mathbf{V},n} \Delta \mathbf{Z}_{\mathbf{V}}) \to 0,$$

as $n \to \infty$, with P-probability one.

Remark 1 (On the regularity of the transportation maps)**.** *The regularity theory of optimal transportation maps is a delicate matter (see [Figalli, 2017]), and is surveyed in [De Philippis and Figalli, 2014]. The classical Cafarelli's regularity theory in [Caffarelli, 1992] requires convexity of the support of the destination measure and that both the source and destination densities be bounded away from* 0 *and infinity. Here, as in [Chernozhukov et al., 2017], a difficulty comes from the fact that the destination density of $P^{\mathbf{S}}$ has a singularity at $\mathbf{0}$, see equation (1.1) in [Figalli, 2018]. Hence, one can not use the classical Cafarelli's regularity, and one has to use the theorems on partial regularity developed in [Figalli and Kim, 2010], [Figalli, 2010], which requires only that the source and destination densities be bounded away from* 0 *and infinity, or the regularity theory in [Figalli, 2018], which sets the source as $P^{\mathbf{S}}$ and requires only boundedness away from zero and infinity for the destination measure. In particular, [Figalli, 2018] gives the missing homeomorphism condition (C) for the validity of Theorem A.2 in [Chernozhukov et al., 2017]. In our setting, another difficulty comes from the fact that if $P^{\mathbf{X}}$ absolutely continuous, it may happen that the source measure $P^{\mathbf{U}}$ has an infinite or zero density on points in the frontier of its support, as e.g. is the case for a Gaussian copula where the density is unbounded at $\mathbf{0}$ and $\mathbf{1}$. Hence, we use the restriction argument to obtain a copula density bounded away from zero and infinity on $\mathbf{L}$ (and $\mathbf{K}$). Note also that $P^{\mathbf{U}_n}$ is always absolutely continuous, since the randomizer $\mathbf{V} \sim \lambda^d$. For the same reason, if $P^{\mathbf{X}}$ is discrete, then $P^{\mathbf{U}}$ has the same properties as those of $P^{\mathbf{U}_n}$.*

3.3 Connection with the depth functions z, z_n of Section 2

The membership depth functions z, z_n have a natural connection with these randomised sets $\mathbf{Z}_{\mathbf{V}}, \mathbf{Z}_{\mathbf{V},n}$, as their conditional expectations:

$$z(\mathbf{x}) = (\mathcal{D}\mathbb{1}_{\mathbf{A}})(\mathbf{x}) = P^*(\mathbf{G}(\mathbf{x}, \mathbf{V}) \in \mathbf{A}) = E^*[\mathbb{1}_{\mathbf{Z_V}}(\mathbf{x})] = E^*[\mathbb{1}_{\mathbf{Z_V}}(\mathbf{X})|\mathbf{X} = \mathbf{x}]$$

and similarly for z_n

$$z_n(\mathbf{x}) = (\mathcal{D}_n \mathbb{1}_{\mathbf{A}_n})(\mathbf{x}) = E^*[\mathbb{1}_{\mathbf{Z}_{\mathbf{V},n}}(\mathbf{x})] = E^*[\mathbb{1}_{\mathbf{Z}_{\mathbf{V},n}}(\mathbf{X})|\mathbf{X} = \mathbf{x}].$$

As a direct corollary, one obtains a simple proof of the coverage probabilities of the membership functions and of the asymptotic nullity of the $L_1(P^{\mathbf{X}})$ distance z and z_n:

Corollary 3.2. *1.* $P^{\mathbf{X}}(z) = P^{\mathbf{X}_n^*}(z_n) = \tau$.

2. With P-probability one, as $n \to \infty$,

$$P^{\mathbf{X}}|z_n - z| \to 0.$$

Proof. 1. By the law of total expectation, $P^{\mathbf{X}}(z) = E^* z(\mathbf{X}) = P^*(\mathbf{X} \in Z_{\mathbf{V}}) = \tau$, and similarly for z_n.

2. By the elementary properties of conditional expectation,

$$\begin{aligned} P^{\mathbf{X}}|z_n - z| &= E^*|z_n(\mathbf{X}) - z(\mathbf{X})| = E^*|E^*[\mathbb{1}_{\mathbf{Z}_{\mathbf{V},n}}(\mathbf{X}) - \mathbb{1}_{\mathbf{Z}_{\mathbf{V}}}(\mathbf{X})|\mathbf{X}]| \\ &\leq E^*\left(E^*[|\mathbb{1}_{\mathbf{Z}_{\mathbf{V},n}}(\mathbf{X}) - \mathbb{1}_{\mathbf{Z}_{\mathbf{V}}}(\mathbf{X})||\mathbf{X}]\right) \\ &= E^*[|\mathbb{1}_{\mathbf{Z}_{\mathbf{V},n}}(\mathbf{X}) - \mathbb{1}_{\mathbf{Z}_{\mathbf{V}}}(\mathbf{X})|] \\ &= P^{\mathbf{X}}(\mathbf{Z}_{\mathbf{V},n} \Delta \mathbf{Z}_{\mathbf{V}}) \to 0 \end{aligned}$$

by Theorem 3.1.

□

4 Empirical depth areas obtained from a smoothed empirical measure

4.1 Motivation and setting

The previous sections showed how to circumvent the complications induced by the discreteness of the empirical measure in order to obtain empirical membership functions z_n, which can be thought of as a set in the enlarged space $\mathbb{R}^d \times [0,1]^d$, or empirical random depth sets $\mathbf{Z}_{\mathbf{V},n}$. If $P^{\mathbf{X}}$ is continuous, $\mathbf{Z}_{\mathbf{V},n}$ is a cloud of the sample points, whereas the population depth set $\mathbf{Z}$ is a "continuum" of $\mathbb{R}^d$. Similarly for the (enlarged) set interpretation of the function z_n. One may consider that this renders these proposals not visually appealing as depth region.

In that regard and in view of the discussion of Section 2, one is naturally inclined in the continuous case to consider a smoothing of the empirical measure in order to obtain a continuous empirical measure. The corresponding transformations of measures are then induced by mapping, i.e. degenerate Markov morphisms, and sets are now transformed into sets.

A probabilistic description of the (kernel) smoothing procedure is as follows: on $(\Omega^*, \mathcal{A}^*, P^*)$ where the r.v.s. live, add to the bootstrap representer $\mathbf{X}_n^*$ of

the empirical measure a small "scaled error" with multivariate bandwidth $\mathbf{h}_n$ from some fixed independent r.v. $\mathbf{W}$ with continuous distribution function K, i.e. define

$$\hat{\mathbf{X}}_n^* := \mathbf{X}_n^* + \mathbf{h}_n \mathbf{W}. \tag{21}$$

The law of $\hat{\mathbf{X}}_n^*$ is the convolution of the empirical measure with the law of $\mathbf{h}_n\mathbf{W}$, i.e.

$$P^{\hat{\mathbf{X}}_n^*} = P^{\mathbf{X}_n^*} * P^{\mathbf{h}_n \mathbf{W}}.$$

Denote by $\hat{F}_n, \hat{\mathbf{G}}_n$ the corresponding joint and marginal (continuous) c.d.f.s. of $\hat{\mathbf{X}}_n^*$. $\hat{F}_n$ corresponds to the well-known kernel smoothed empirical cdf,

$$\hat{F}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{\mathbf{h}_n} \right),$$

where K is the joint cdf of $\mathbf{W}$. Since $\hat{F}_n$ is continuous, one can define the empirical copula representer $\hat{\mathbf{U}}_n$ via the Multivariate Marginal Probability Integral Transform,

$$\hat{\mathbf{U}}_n := \hat{\mathbf{G}}_n(\hat{\mathbf{X}}_n^*), \tag{22}$$

and denote by $\hat{C}_n$ its (copula) c.d.f. The rest of the procedure is as before: Monge-mass transport $P^{\hat{\mathbf{U}}_n}$ to $P^{\mathbf{S}}$ by the transport map $\mathbf{R}_{\hat{C}_n}$ with inverse $\mathbf{Q}_{\hat{C}_n}$. Eventually, $\hat{\mathbf{S}}_n$ is obtained by setting $\hat{\mathbf{S}}_n = \mathbf{R}_{\hat{C}_n}(\hat{\mathbf{U}}_n)$. One has transformed all corresponding measures by push-forwarding them by mappings and so we can reason at the level of random variables according to the diagram in Figure 6.

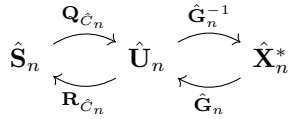

Figure 6: Transformations of the smooth empirical measure by mappings of the corresponding random variables

For a continuous $P^{\mathbf{X}}$, the distributional transform $\mathbf{X} \to \mathbf{G}(\mathbf{X}, \mathbf{V})$ reduces to the probability integral transform $\mathbf{X} \to \mathbf{G}(\mathbf{X})$, so the population counterparts of (21) and (22) are obtained similarly by transformations of random variables by non-randomized mappings, as in the diagram in Figure 7. (Recall that $\mathbf{S}, \mathbf{U}, \mathbf{X}$ are defined on $(\Omega^*, \mathcal{A}^*, P^*)$).

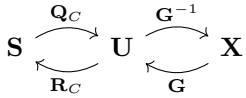

Figure 7: Transformations of random variables–population version for a continuous $\mathbf{X}$

One obtains for both the empirical and the population depths areas at the $\mathbf{X}$ level genuine sets $\mathbf{Z}, \hat{\mathbf{Z}}_n$, defined naturally as

$$\mathbf{Z} = \mathbf{G}^{-1}(\mathbf{Q}_C(\mathbf{B})), \quad \hat{\mathbf{Z}}_n = \hat{\mathbf{G}}_n^{-1}(\mathbf{Q}_{\hat{C}_n}(\mathbf{B})).$$

Remark 2 (On bandwidth and kernel choice)**.** *In (21) it is recommended to choose* $\mathbf{W}$ *with independent components, i.e. a product kernel* K*, so that one does not introduce artificial dependence in the components of* $\hat{\mathbf{X}}_n^*$*. Moreover,* $\hat{\mathbf{U}}_n$ *in (22) is obtained from* $\hat{\mathbf{X}}_n^*$ *by the transform* $\hat{\mathbf{G}}_n$ *which acts marginal by marginal. Therefore, for the choice of the multivariate bandwidth* $\mathbf{h}_n$*, one can use univariate bandwidth choice techniques for each component of* $\hat{\mathbf{G}}_n$ *and optimize each marginal bandwidth of* $\mathbf{h}_n$ *separately.*

4.2 Asymptotic results

The following Proposition is the analogue of the main Theorem 6.2 in [Faugeras and Rüschendorf, 2017]:

Proposition 4.1. *If* $P^{\mathbf{X}}$ *is absolutely continuous, and* $\mathbf{h}_n \downarrow \mathbf{0}$*, one has, with* P *probability one,*

$$(\hat{\mathbf{X}}_n^*, \hat{\mathbf{U}}_n, \hat{\mathbf{S}}_n) \overset{P^* a.s.}{\longrightarrow} (\mathbf{X}, \mathbf{U}, \mathbf{S}).$$

In turn, Proposition 4.1 translates at the level of depth sets as follows:

Corollary 4.2. *With* $P-$*probability one, if* $P^{\mathbf{X}}$ *is continuous, the* $P^{\mathbf{X}}$ *symmetric distance between the population depth area* $\mathbf{Z}$ *and its empirical smoothed counterpart* $\hat{\mathbf{Z}}_n$ *becomes asymptotically negligible,*

$$P^{\mathbf{X}}(\hat{\mathbf{Z}}_n \Delta \mathbf{Z}) \to 0,$$

so that $\hat{\mathbf{Z}}_n$ *is asymptotically of* $P^{\mathbf{X}}$*-mass* τ*:* $P^{\mathbf{X}}(\hat{\mathbf{Z}}_n) \to \tau$*.*

Proof. Similar to the proof of Theorem 2.1 1. (c) and 2. (b). □

5 Simulations

In this section, we provide some extensive numerical illustrations of the proposed multivariate quantile and depth areas, illustrate their properties and compare them with results of [Chernozhukov et al., 2017] and [del Barrio et al., 2018].

5.1 Basic algorithm and numerical implementations of optimal transport

The most crucial step for obtaining the empirical quantile areas is the computation of the optimal transport map between the reference spherical distribution $P^{\mathbf{S}}$ on the unit ball and the empirical copula distribution $P^{\mathbf{U}_n}$ (or $P^{\hat{\mathbf{U}}_n}$). Numerical implementation of optimal transport is a subject of active research, see

[Peyré et al., 2019] for a survey of available methods. Optimal transportation between discrete distributions reduces to a finite dimensional linear program, which further reduces to an optimal matching problem when the mass of each Dirac is constant and the two measures have the same number of Dirac masses, see e.g. [Peyré et al., 2019] Chapters 2 and 3. Therefore, and also for comparison purposes with [Chernozhukov et al., 2017] and [del Barrio et al., 2018], we chose to implement algorithms of optimal transport for discrete to discrete distributions. This requires to discretize $P^{\mathbf{S}}$ and $P^{\mathbf{U}_n}$ (or $P^{\hat{\mathbf{U}}_n}$), and is done as explained thereafter.

The general procedure to compute the proposed randomized depth areas of Section 3 is described as follows:

1. Sample n data points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from the $P^{\mathbf{X}}$ distribution to be analyzed.

2. Discretize the empirical copula measure as follows: Since $P^{\mathbf{U}_n}$ is the law of $\mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V})$, with $\mathbf{X}_n^* \sim F_n$, one can simply obtain its discretization by generating n i.i.d. realizations $\mathbf{V}_1, \ldots, \mathbf{V}_n$ of the randomizer $\mathbf{V}$, uniformly distributed on $[0,1]^d$, and computing the empirical copula points as $\mathbf{G}_n(\mathbf{X}_i, \mathbf{V}_i)$, for $i = 1, \ldots, n$.

   Note also that for large sample size, one can speed up the computations by bypassing this distributional transform step, and compute an approximate empirical copula sample by simply taking $\mathbf{G}_n(\mathbf{X}_i)$, for $i = 1, \ldots, n$ as empirical copula points.

3. Discretize the unit ball where $P^{\mathbf{S}}$ lives into a regular grid of n points.

4. Compute the optimal transportation matrix from the n points of the discretized ball into the n empirical copula points, and deduce the corresponding optimal transportation map $\mathbf{Q}_{C_n}$.

5. Generate the quantile areas/contours from the unit ball by picking the points inside/on the sphere $\mathbf{B}_\tau$ of radius τ. Obtain the empirical quantile areas/contours $\mathbf{A}_n = \mathbf{Q}_{C_n}(\mathbf{B}_\tau)$ in the copula space $[0,1]^d$ by mapping those points with the optimal transportation map of step 4.

6. Obtain the empirical quantile areas/contours $\mathbf{Z}_n = \mathbf{G}_n^{-1}(\mathbf{A}_n)$ in the sample space $\mathbb{R}^d$ of $P^{\mathbf{X}}$ by mapping the quantile contour points $\mathbf{A}_n$ in the copula space $[0,1]^d$ with the marginal empirical quantile transform $\mathbf{G}_n^{-1}$, computed from the sample of step 1.

For the smoothed version of Section 4, the algorithm is similar, with using the smoothed version $\hat{F}_n$, $\hat{\mathbf{G}}_n$ instead of the empirical measure in steps 2 and 6. Note also that one does not need randomizers in this case, see (22).

We implemented the linear programming version of optimal transportation, as in [del Barrio et al., 2018]. However, it runs into memory trouble for sample size *nboot* exceeding 800 points on our computer (the codes were written in Mathematica, on a computer with 8Gb RAM). Therefore, we also implemented the entropy regularised optimal transportation method introduced by [Cuturi,

2013]. Although it gives only an approximate optimal transport matrix, the computations are extremely fast, even for large sample size, and is thus our recommended method. In addition, two variants (the mode method and the barycenter method) are proposed in the Sinkhorn algorithm for the computation of the optimal transportation map from the transportation plan.

Details on the implementation and comparison of algorithms are presented in Appendix 7. The randomized depth areas approach of Section 3 and the smoothed approach of Section 4 give similar results, with the smoothing approach yielding the most visually appealing graphics, while the randomized depth areas approach is slightly faster. The implementation by the Sinkhorn algorithm + barycenter method works well for distributions with "nice", convex, connected support, while the Sinkhorn algorithm + mode method, by mapping to an exact point of the sample which therefore will fall inside the support and respect its geometry, should be preferred otherwise. See Appendix 7 for the definitions of the methods considered.

5.2 Some numerical illustrations

We present below some numerical illustrations of the properties of the quantile areas obtained. For obvious graphical reasons, we only consider bivariate models.

5.2.1 Distributions with non-convex or non-connected support

In order to illustrate the ability of the proposed approach to pick the correct geometry of a distribution with non-convex support, we simulated $n = 5000$ observations from the regression model $Y = -1 + X^2 + \epsilon$, with $X \sim U_{[-2,2]}$ independent of $\epsilon \sim U_{[-1,1]}$, see Figure 8. By construction, the distribution of $\mathbf{X} := (X, Y)$ has a non-convex support: it is a band curved around a parabola. Such a distribution has a "banana" shape similar to the one in Figures 1 and 2 in Chernozhukov et al. [2017].

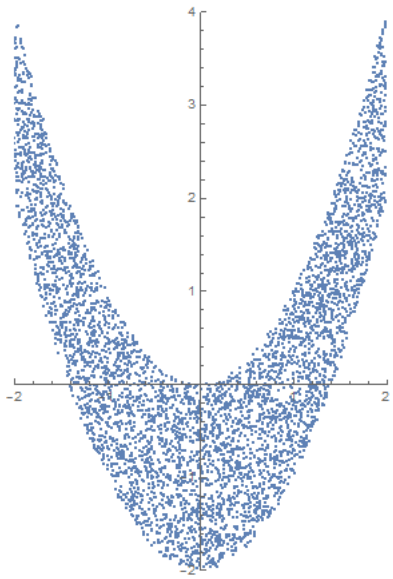

Figure 8: Scatterplot of $n = 5000$ i.i.d. observations from the regression model $\mathbf{X} = (X, Y)$, $Y = -1 + X^2 + \epsilon$, with $X \sim U_{[-2,2]}$ independent of $\epsilon \sim U_{[-1,1]}$.

In Figure 9, we displayed the $\tau = 0.5$ quantile region obtained using Sinkhorn's algorithm with extraction of the transportation map by the mode method in the observation space. The observed sample points which fall inside the $0.5-$quantile region are represented by red circles, and those outside by blue filled squares: the half central quantile area nicely adapts to the non-convex "banana" geometry of the distribution considered.

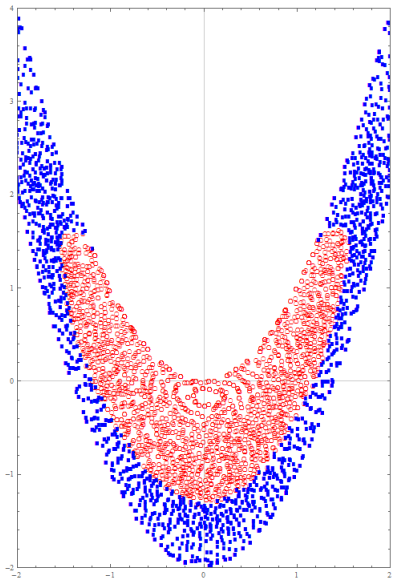

Figure 9: $\tau = 0.5$ non-convex quantile region obtained using Sinkhorn's algorithm with extraction of the transportation map by the mode method. (Red) circles: sample points inside the $0.5-$quantile region, (Blue) filled squares: sample points sample points outside the $0.5-$quantile region.

Analogously, we illustrate the ability of the proposed method to pick non-connected quantile areas in Figure 10: we simulated $n = 5000$ observations from a distribution uniformly distributed on two disjoint unit disks. By construction, the distribution has a disconnected support. The $\tau = 0.5$ quantile region obtained using the same implementation of the algorithm as above is displayed by the points marked by red circles: the quantile region is also a non-connected central area in the disjoint support of the two disks.

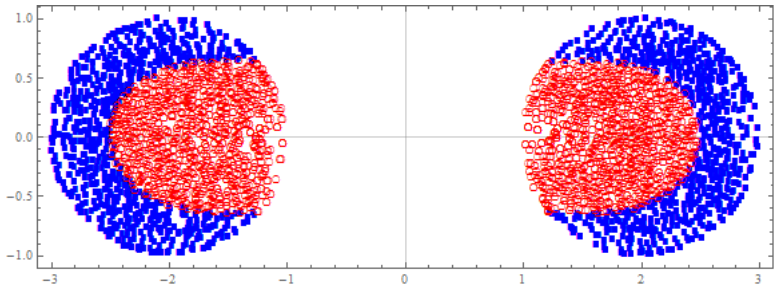

Figure 10: $\tau = 0.5$ non-connected quantile region obtained using Sinkhorn's algorithm with extraction of the transportation map by the mode method. (Red) circles: sample points inside the $0.5-$quantile region, (Blue) filled squares: sample points sample points outside the $0.5-$quantile region.

5.2.2 Convergence of depth areas

In order to illustrate the convergence properties of the empirical quantile areas to their population counterparts as the sample size increases, we take for $\mathbf{X}$ a standard bivariate Gaussian distribution, whose theoretical quantile areas are known and easy to calculate: they are disks whose radius is the corresponding quantile of the Rayleigh distribution.

Figure 11 illustrates Corollaries 4.2, 3.2 of the smoothed approach of Section 4). We have drawn the empirical quantile contours for $\tau = .25, .50, .75, .90$ (colored lines) together with their theoretical counterparts (shaded disks) and sample data clouds. Compared to Figure 2.1 in [del Barrio et al., 2018], one also obtains nicely converging nested regions, as expected, somehow slightly less "spiked". The unsmoothed approach of Section 3 gives a similar picture, and is therefore omitted. Notice that for high quantile areas ($\tau = 0.9$), the convergence takes longer to occur, as one enters the domain of application of extreme value theory.

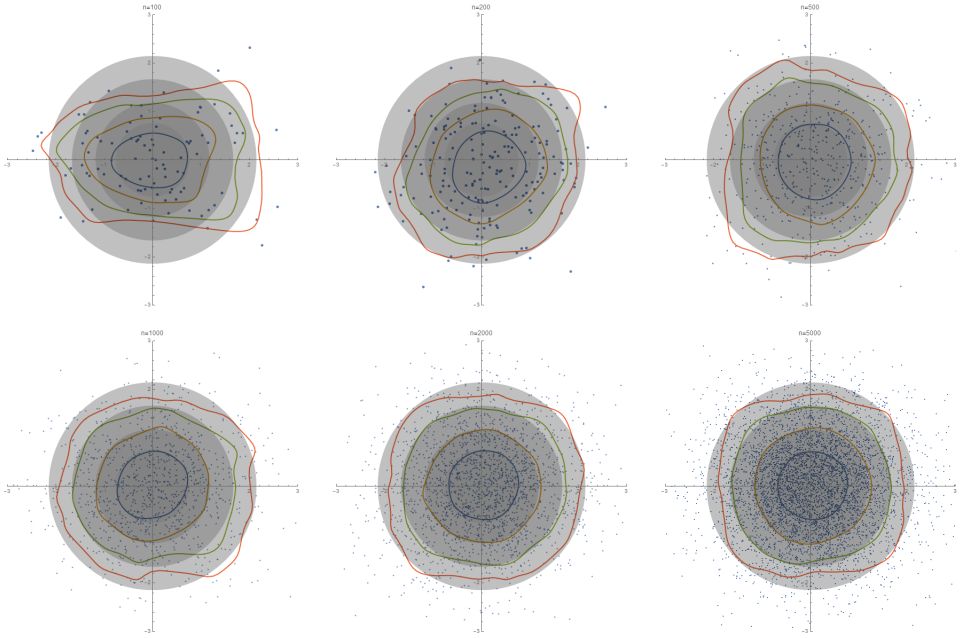

Figure 11: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 (green), .90 (red)) computed from $n = 100$, 200, 500, 1000, 2000, 5000 i.i.d. observations from a bivariate standard Gaussian distribution, along with their (spherical) theoretical counterparts (shaded disks). (Sinkhorn algorithm + barycenter method)

5.2.3 Monotone invariance

For illustrating the monotone equivariance property of the proposed quantile areas w.r.t. monotone transformations of the marginals, we drew the empirical quantile contours in the copula space $[0,1]^d$ and in the original sample space $\mathbb{R}^d$ for a Frank copula model with varied marginals (smoothed approach).

In Figure 12, we simulated $n = 1000$ sample points from a Frank (with parameter $\theta = 500$) copula distribution with uniform marginals. As expected with uniform marginals, we obtain the same quantile contours in the copula space $[0,1]^2$ (left panel) as in the sample space $\mathbb{R}^2$ (right panel).

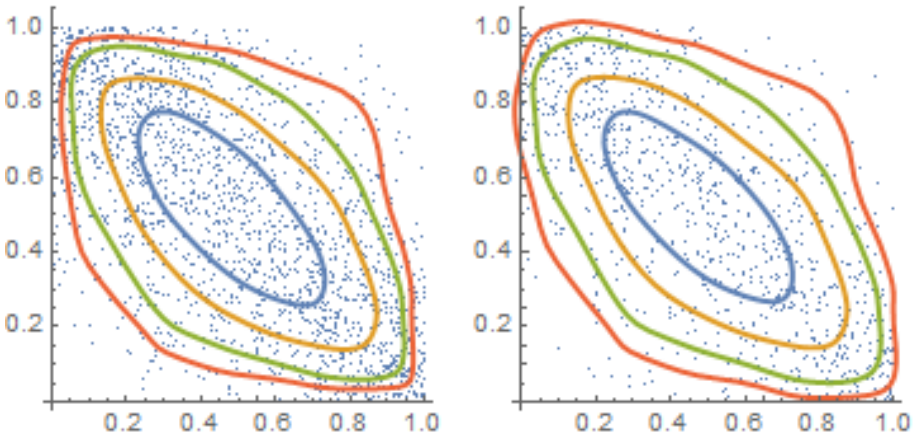

Figure 12: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 green), .90 red)) computed from $n = 1000$ i.i.d. observations from a Frank ($\theta = 500$) copula distribution with Uniform marginals. The left panel shows the empirical contours in the copula space together with the bootstrap sample from the empirical copula. The right panel shows the empirical contours in the original sample space together with the observed data. (Sinkhorn algorithm + barycenter method)

In Figure 13, we simulated $n = 1000$ sample points from the same Frank copula, but with both marginals changed to two Exponential $Exp(2)$ distributions. In the copula space (left panel), one obtains, up to sample fluctuations, the same contour regions as in Figure 12, which is to be expected since copulas are invariant w.r.t monotone increasing marginal transformations. However, in the original sample space, one now obtains empirical quantile regions stretched by the Exponential marginal transformation, which gives depth areas nicely located in the concentration areas of the data.

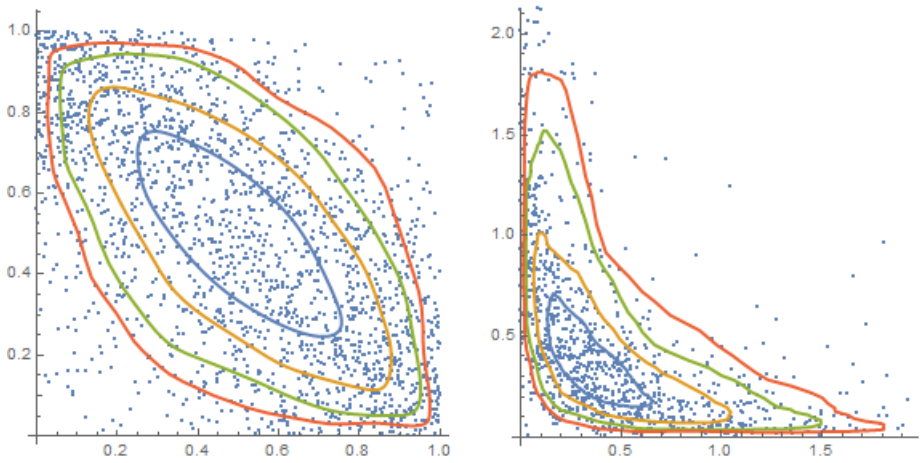

Figure 13: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 green), .90 red)) computed from $n = 1000$ i.i.d. observations from a Frank ($\theta = 500$) copula distribution with Exponentials $Exp(2)$ marginals. The left panel shows the empirical contours in the copula space together with the bootstrap sample from the empirical copula. The right panel shows the empirical contours in the original sample space together with the observed data. (Sinkhorn algorithm + barycenter method)

5.2.4 Gaussian Mixtures

For comparison sake, we also plotted the empirical contours obtained for some of the mixture distributions studied in [del Barrio et al., 2018], Section 2.3.2. Figure 14 is the analogue of Figure 2.2 in [del Barrio et al., 2018] for the symmetric Gaussian mixture distributions $\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-1\\0\end{pmatrix}, I_2\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix}1\\0\end{pmatrix}, I_2\right)$ (left panel), $\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-2\\0\end{pmatrix}, I_2\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix}2\\0\end{pmatrix}, I_2\right)$ (center panel), $\frac{1}{2}\mathcal{N}\left(\begin{pmatrix}-4\\0\end{pmatrix}, I_2\right) + \frac{1}{2}\mathcal{N}\left(\begin{pmatrix}4\\0\end{pmatrix}, I_2\right)$ (right panel). We obtain a similar shape of the quantile contours as in del Barrio et al. [2018], slightly more regular. As in [del Barrio et al., 2018], we emphasize the difference between quantile regions and level-sets of the density: here, the quantile regions obtained in this example are central, connected and nested areas, whereas the density level sets would give disconnected regions separating the two modes of the mixture.

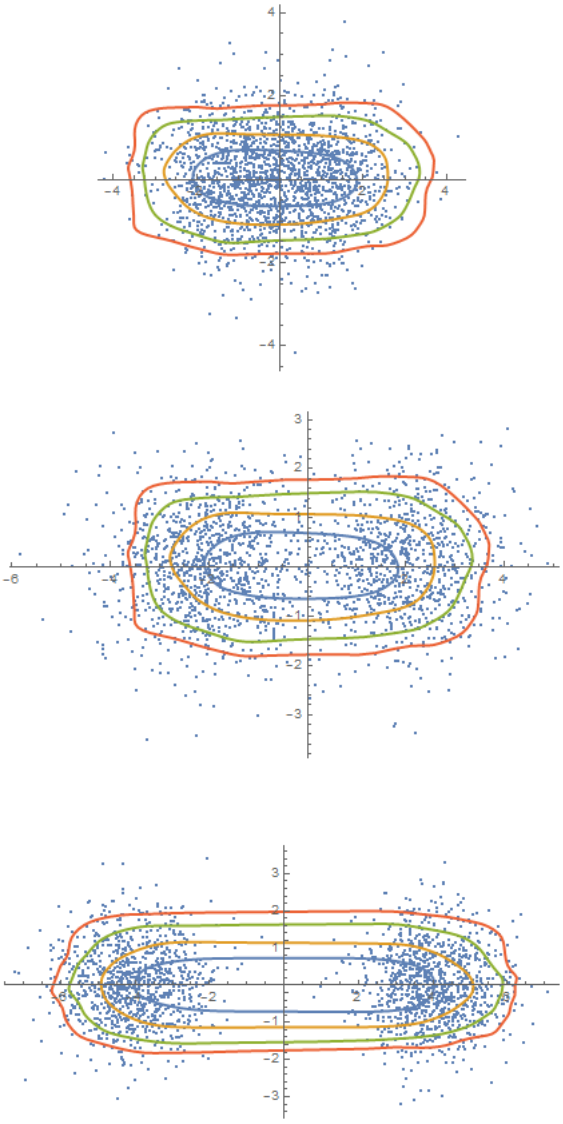

Figure 14: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 (green), .90 (red)) computed from $n = 2000$ i.i.d. observations from symmetric Gaussian mixture distributions .

Conclusion

We proposed three multivariate notions of central quantile regions: membership functions z, randomized depth areas $\mathbf{Z}_{\mathbf{V}}$, and marginally smoothed depth areas $\hat{\mathbf{Z}}$. These are based on the copula and mass transportation approach of [Faugeras and Rüschendorf, 2017] to multivariate quantiles. In particular, their empirical counterpart z_n, $\mathbf{Z}_{\mathbf{V},n}$, $\hat{\mathbf{Z}}_n$ attain the exact level and are also strongly consistent. These results correct Corollary 6.4 in [Faugeras and Rüschendorf, 2017], and improve on those of [Chernozhukov et al., 2017]. Simulations illustrate how these "depth areas" can be numerically obtained and show that they give a valid representation of the central quantile area of a multivariate distribution, and thus are a valuable tool for their analysis. Let us stress the interest of our proposed approach and summarize the results obtained:

- The method of a single probability space we used in the proofs, see (10), allows to obtain convergence $P-$a.s., and not only in P-probability as in [Chernozhukov et al., 2017]. Also, let us stress that the copula step allows to reduce the optimal transportation problem to measures on bounded domains, and so one does not need assumptions on the existence of second moment of $P^{\mathbf{X}}$, nor compactness of its support as in [Chernozhukov et al., 2017].

- As may appear surprising at first sight, the Markov morphism view on multivariate quantiles we advocate in [Faugeras and Rüschendorf, 2017] and in this paper allows to define multivariate quantile objects even for a discrete distribution $P^{\mathbf{X}}$! The price to be paid is a conceptual change in the corresponding notion of "multivariate central quantile areas": one either has to introduce membership *functions* z, (which can be interpreted as indicators of a genuine set, but in an enlarged space), or consider *randomized* quantile sets $\mathbf{Z}_{\mathbf{V}}$. This is also convenient for their statistical counterparts z_n and $\mathbf{Z}_{\mathbf{V},n}$, where one must make inference from the discrete empirical measure $P^{\mathbf{X}_n^*}$.

- The copula approach based on the distributional transform allows to obtain genuine absolutely continuous empirical copulas, see [Faugeras, 2015], [Faugeras, 2017]. This results in a nonparametric estimation procedure without any bandwidth to optimize. In particular, the empirical membership function z_n and randomized depth area $\mathbf{Z}_{\mathbf{V},n}$ are bandwidth free. For the marginally smoothed quantile areas $\hat{\mathbf{Z}}_n$, one simply has d-univariate bandwidths to optimize in the kernel smoothing step, which is much easier than smoothing a d-variate distribution, as in [Chernozhukov et al., 2017], where one has in practice a $d \times d$ positive definite matrix of bandwidths to optimize. Note also that multivariate density estimation is subject to the curse of dimensionality.

- Because copulas are invariant w.r.t. monotone increasing transformations of the marginals, see e.g. Theorem 2.4.3 in [Nelsen, 2006], one obtains

depth areas which are fully equivariant w.r.t. monotone increasing transformations of the marginals of $\mathbf{X}$. In other words, if $\mathbf{X}$ is transformed into $\mathbf{T}(\mathbf{X}) := (T_1(X_1), \ldots, T_d(X_d))$, where each map $T_1, \ldots, T_d$ are strictly increasing, the depth areas will remain the same at the copula level, and will be transformed at the $\mathbf{X}$ level by the corresponding $\mathbf{T}$. One has therefore a key feature of univariate quantiles, equivariance w.r.t. a nonlinear monotone change of scale, extended to the multivariate setting. This is more general than the affine equivariance of [Chernozhukov et al., 2017]. Note also that requiring equivariance of the quantile region under a multivariate affine transformation $\mathbf{x} \to T\mathbf{x} + \mathbf{b}$, where T a $d \times d$ matrix, as argued in [del Barrio et al., 2018], seems ill-suited from a physical standpoint: if the components of $\mathbf{X}$ stands for quantities expressed in different units (say X_1 =mass in (kg), X_2 =speed in (m.s^{-1}), etc.), it does not make sense to take linear combination of them, as the variables are non-commensurable. To the contrary, the copula transformation step in the present approach encodes the different magnitudes scales into a single, common, dimensionless propensity scale, as marginal quantile and distribution functions are in Galois connections of each other, see Faugeras and Rüschendorf [2017]. Hence, the proposed combined copula and mass transportation approach to multivariate quantile is robust w.r.t. general (nonlinear) changes of scales.

- On the practical and numerical side, the proposed depth areas are easily computable thanks to a fast implementation of the optimal transport step using the entropy regularised / Sinkhorn algorithm. This give a new valuable tool to the applied scientist for the analysis of multivariate data.

References

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991. ISSN 0010-3640. doi: 10.1002/cpa.3160440402. URL `https://doi.org/10.1002/cpa.3160440402`.

Luis A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992. ISSN 08940347, 10886834. URL `http://www.jstor.org/stable/2152752`.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45 (1):223–256, 2017.

J. A. Cuesta-Albertos, C. Matran, and A. Tuero-Diaz. Optimal transportation plans and convergence in distribution. *Journal of Multivariate Analysis*, 60: 72–83, 1997.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems (NIPS 2013)*, volume 26, pages 2292–2300, 2013.

G. De Philippis and A. Figalli. Partial regularity results in optimal transportation. In *Trends in contemporary mathematics*, volume 8 of *Springer INdAM Ser.*, pages 293–307. Springer, Cham, 2014.

E. del Barrio, J.A. Cuesta-Albertos, M. Hallin, and C. Matrán. Center-outward distribution functions, quantiles, ranks, and signs in $\mathbb{R}^d$. *Preprint*, 2018. URL `https://arxiv.org/pdf/1806.01238.pdf`.

W. Edwards Deming and Frederick F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statistics*, 11:427–444, 1940. ISSN 0003-4851. doi: 10.1214/aoms/1177731829. URL `https://doi.org/10.1214/aoms/1177731829`.

Luc Devroye and László Györfi. *Nonparametric density estimation.* Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons, Inc., New York, 1985. ISBN 0-471-81646-9. The L_1 view.

George S. Easton and Robert McCulloch. A multivariate generalization of quantile–quantile plots. *Journal of the American Statistical Association*, 85 (410):376–386, 1990. ISSN 0162-1459. doi: 10.1080/01621459.1990.10476210.

Olivier P. Faugeras. Maximal coupling of empirical copulas for discrete vectors. *Journal of Multivariate Analysis*, 137(11):179–186, 2015.

Olivier P. Faugeras. Inference for copula modeling of discrete data: a cautionary tale and some facts. *Depend. Model.*, 5(1):121–132, 2017. ISSN 2300-2298. doi: 10.1515/demo-2017-0008. URL `https://doi.org/10.1515/demo-2017-0008`.

Olivier P. Faugeras and Ludger Rüschendorf. Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles. *Mathematica Applicanda*, 45(1):3–45, 2017.

Alessio Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35(3):465–479, 2010. ISSN 0360-5302. doi: 10.1080/03605300903307673. URL `https://doi.org/10.1080/03605300903307673`.

Alessio Figalli. *The Monge-Ampère equation and its applications.* Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2017. ISBN 978-3-03719-170-5. doi: 10.4171/170. URL `https://doi.org/10.4171/170`.

Alessio Figalli. On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.*, 177(part B):413–421, 2018. ISSN 0362-546X. doi: 10.1016/j.na.2018.05.008. URL https://doi.org/10.1016/j.na.2018.05.008.

Alessio Figalli and Young-Heon Kim. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Contin. Dyn. Syst.*, 28(2): 559–565, 2010. ISSN 1078-0947. doi: 10.3934/dcds.2010.28.559. URL https://doi.org/10.3934/dcds.2010.28.559.

Marc Hallin. On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: a measure transportation approach. *Preprint*, 2017. URL https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995. ISSN 0012-7094. doi: 10.1215/S0012-7094-95-08013-2. URL https://doi.org/10.1215/S0012-7094-95-08013-2.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, 2006.

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Georg Pólya. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.*, 8(3-4):171–181, 1920. ISSN 0025-5874. doi: 10.1007/BF01206525. URL https://doi.org/10.1007/BF01206525.

R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2):659–680, 1962.

Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987. ISBN 0-07-054234-1.

L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1):48–54, 1990. ISSN 0047-259X. doi: 10.1016/0047-259X(90)90070-X. URL https://doi.org/10.1016/0047-259X(90)90070-X.

R. S. Singh, T. Gasser, and B. Prasad. Nonparametric estimates of distribution functions. *Comm. Statist. A—Theory Methods*, 12(18):2095–2108, 1983. ISSN 0361-0926. doi: 10.1080/03610928308828593. URL https://doi.org/10.1080/03610928308828593.

Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964. ISSN 0003-4851. doi: 10.1214/aoms/1177703591. URL https://doi.org/10.1214/aoms/1177703591.

Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly*, 74:402–405, 1967. ISSN 0002-9890. doi: 10.2307/2314570. URL https://doi.org/10.2307/2314570.

Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, 21:343–348, 1967. ISSN 0030-8730. URL http://projecteuclid.org/euclid.pjm/1102992505.

C. S. Smith and M. Knott. Note on the optimal transportation of distributions. *J. Optim. Theory Appl.*, 52(2):323–329, 1987. ISSN 0022-3239. doi: 10.1007/BF00941290. URL https://doi.org/10.1007/BF00941290.

Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. doi: 10.1007/978-1-4757-2545-2. URL https://doi.org/10.1007/978-1-4757-2545-2. With applications to statistics.

B. B. Winter. Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.*, 1(2):247–253, 1973. ISSN 0319-5724. doi: 10.2307/3315003. URL https://doi.org/10.2307/3315003.

Hajime Yamato. Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, 15(3-4):69–78, 1972/73.

J. E. Yukich. A note on limit theorems for perturbed empirical processes. *Stochastic Process. Appl.*, 33(1):163–173, 1989. ISSN 0304-4149. doi: 10.1016/0304-4149(89)90073-2. URL https://doi.org/10.1016/0304-4149(89)90073-2.

6 Appendix 1: Proofs

6.1 Absolute continuity of copula measures of absolutely continuous distributions

If F is continuous, the copula is unique and $P^{\mathbf{U}}$ is the push-forward of $P^{\mathbf{X}}$ by $\mathbf{G}$. If, in addition, $P^{\mathbf{X}}$ is absolutely continuous, it is not immediately clear that $P^{\mathbf{U}}$ is also absolutely continuous, since $\mathbf{G}$ may have flat spots (the result would be obvious if the marginals of $\mathbf{X}$ had strictly positive densities). Intuitively, the points were the components of $\mathbf{G}$ have a zero derivative are those which are not charged by $P^{\mathbf{X}}$. The following lemma on the absolute continuity of the copula measure gives a proof of this statement and is of independent interest.

Lemma 6.1. *If $P^{\mathbf{X}} \ll \lambda^d$, then $P^{\mathbf{U}} \ll \lambda^d$.*

Proof. The proof is based on the transformation formula Theorem 7.26 in Rudin [1987] p. 154: for every measurable function $h : \mathbb{R}^d \mapsto [0, \infty]$, one has

$$\int_{\mathbf{T}(\mathbf{N})} h d\lambda^d = \int_{\mathbf{N}} (h \circ \mathbf{T}) |J_{\mathbf{T}}| d\lambda^d, \tag{23}$$

for a Lebesgue measurable set $\mathbf{N} \subset \mathbf{O} \subset \mathbb{R}^d$, s.t. $\mathbf{O}$ is open, $\mathbf{T} : \mathbf{O} \mapsto \mathbb{R}^d$ is continuous, $\mathbf{T}$ differentiable and one-to-one on $\mathbf{N}$, and $\lambda^d(\mathbf{T}(\mathbf{O} \setminus \mathbf{N})) = 0$.

Let $\mathbf{G}(\mathbf{x}) = (G_1(x_1), \ldots, G_d(x_d))$ be the probability integral transform of $\mathbf{X} \sim F$, with marginals $\mathbf{G}$. Let f denotes the density of $P^{\mathbf{X}} = f\lambda^d$, with g_i the density of the ith component of f. Let $\mathbf{A} \in \mathcal{B}([0,1]^d)$ be a Borel-measurable set s.t. $\lambda^d(\mathbf{A}) = 0$.

Let $\mathbf{E} = \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) > 0\}$ and $\mathbf{M} = \overbrace{\prod \{x_i \in \mathbb{R} : g_i(x_i) > 0\}}^{\circ}$ be the open interior. If $\mathbf{x} \in \mathbf{M}^c$, then $\exists i \in \{1, \ldots, d\}$ s.t. $0 = g_i(x_i) = \int f(\mathbf{x}) d\lambda^{d-1}_{-i}$, where the integration is w.r.t. all coordinates $x_1, \ldots, x_d$, except x_i. Therefore, $f(\mathbf{x}) = 0$ a.e. Hence, $\mathbf{E} \subset \mathbf{M}$ a.e.

Thus, one has

$$\begin{aligned} P^{\mathbf{U}}(\mathbf{A}) &= P^{\mathbf{G}(\mathbf{X})}(\mathbf{A}) = \int \mathbb{1}_{\mathbf{A}}(\mathbf{G}(\mathbf{x})) P^{\mathbf{X}}(d\mathbf{x}) \\ &= \int \mathbb{1}_{\mathbf{A}}(\mathbf{G}(\mathbf{x})) \mathbb{1}_{\mathbf{E}}(\mathbf{x}) f(\mathbf{x})(d\mathbf{x}) = \int \mathbb{1}_{\mathbf{A}}(\mathbf{G}(\mathbf{x})) \mathbb{1}_{\mathbf{E} \cap \mathbf{M}}(\mathbf{x}) f(\mathbf{x})(d\mathbf{x}). \end{aligned}$$

On $\mathbf{G}(\mathbf{M})$, $\mathbf{G}^{-1}$ is differentiable, one-to-one, with inverse $\mathbf{G}$ continuous. Moreover, the Jacobian of $\mathbf{G}^{-1}$ is well defined on $\mathbf{G}(\mathbf{M})$ and writes as $|J_{\mathbf{G}^{-1}}(\mathbf{u})| = \left(\prod g_i(G_i^{-1}(u_i))\right)^{-1}$. Therefore, one has that

$$\begin{aligned} P^{\mathbf{U}}(\mathbf{A}) &= \int \mathbb{1}_{\mathbf{G}^{-1}(\mathbf{A}) \cap \mathbf{G}^{-1}(\mathbf{G}(\mathbf{E} \cap \mathbf{M}))}(\mathbf{x}) f(\mathbf{x})(d\mathbf{x}) \\ &= \int_{\mathbf{G}^{-1}(\mathbf{A} \cap \mathbf{G}(\mathbf{E} \cap \mathbf{M}))} f(\mathbf{x})(d\mathbf{x}) \\ &= \int_{\mathbf{A} \cap \mathbf{G}(\mathbf{E} \cap \mathbf{M})} \frac{f(\mathbf{G}^{-1}(\mathbf{u}))}{\prod g_i(G_i^{-1}(u_i))} d\mathbf{u}, \end{aligned} \tag{24}$$

where (24) follows, by regularity of the Lebesgue measure, from (23) applied to $\mathbf{T} := \mathbf{G}^{-1}$ and $\mathbf{N} := \mathbf{A} \cap \mathbf{G}(\mathbf{E} \cap \mathbf{M})$.

For any $L > 0$, let

$$\mathbf{L} := \{\mathbf{u} \in [0,1]^d : f(\mathbf{G}^{-1}(\mathbf{u})) \leq L, \text{ and } \prod g_i(G_i^{-1}(u_i)) \geq 1/L\}.$$

On $\mathbf{G}(\mathbf{E} \cap \mathbf{M})$, $0 < \frac{f(\mathbf{G}^{-1}(\mathbf{u}))}{\prod g_i(G_i^{-1}(u_i))} < \infty$, hence $\mathbf{L} \subset \mathbf{G}(\mathbf{E} \cap \mathbf{M})$. Thus, by (24), one has that $P^{\mathbf{U}}(\mathbf{A} \cap \mathbf{L}) \leq \lambda^d(\mathbf{A})$. By assumption, $\lambda^d(\mathbf{A}) = 0$, hence $P^{\mathbf{U}}(\mathbf{A} \cap \mathbf{L}) = 0$, for all $L > 0$. By monotone convergence, this entails $P^{\mathbf{U}}(\mathbf{A}) = 0$, i.e. $P^{\mathbf{U}} \ll \lambda^d$. □

6.2 Proofs of the main results

6.2.1 Proof of Theorem 2.1

Proof. 1. (a) follows from the definitions of the Markov morphisms:

$$P^{\mathbf{U}}(\mathbf{A}) = P^{\mathbf{U}}(a) = P^{\mathbf{U}}(\mathcal{R}_C b) = (P^{\mathbf{U}}\mathcal{R}_C)(b) = P^{\mathbf{S}}(b) = \tau,$$

and similarly

$$P^{\mathbf{U}_n}(\mathbf{A}_n) = P^{\mathbf{U}_n}(a_n) = P^{\mathbf{U}_n}(\mathcal{R}_{C_n} b) = (P^{\mathbf{U}_n}\mathcal{R}_{C_n})(b) = P^{\mathbf{S}}(b) = \tau.$$

(b) $|P^{\mathbf{U}}(a_n - a)| \leq P^{\mathbf{U}}|a_n - a| \to 0$, as $n \to \infty$ by Theorem 2.1 1. (c).

(c) One has

$$\begin{aligned} P^{\mathbf{U}}(|a_n - a|) &= P^{\mathbf{U}}(|\mathbb{1}_{\mathbf{A}_n} - \mathbb{1}_{\mathbf{A}}|) \\ &= P^*(\mathbf{U} \in A, \mathbf{U} \notin A_n) + P^*(\mathbf{U} \notin A, \mathbf{U} \in A_n) \\ &:= (I) + (II) \end{aligned}$$

with

$$(I) = P^*(\mathbf{U} \in \mathbf{A}, \mathbf{U} \notin \mathbf{A}_n, \mathbf{U}_n \in \mathbf{A}_n) + P^*(\mathbf{U} \in \mathbf{A}, \mathbf{U} \notin \mathbf{A}_n, \mathbf{U}_n \notin \mathbf{A}_n),$$

where $\mathbf{U}, \mathbf{U}_n$ are the coupling constructions of Theorem 6.2 in [Faugeras and Rüschendorf, 2017]. Let d be the Euclidean distance on $\mathbb{R}^d$. By the triangle inequality,

$$d(\mathbf{U}, \mathbf{A}_n) \leq d(\mathbf{U}, \mathbf{U}_n) + d(\mathbf{U}_n, \mathbf{A}_n)$$

By Theorem 6.2 in [Faugeras and Rüschendorf, 2017], $d(\mathbf{U}, \mathbf{U}_n) \overset{P^* a.s.}{\longrightarrow} 0$, and if $\mathbf{U}_n \in \mathbf{A}_n$ then $d(\mathbf{U}_n, \mathbf{A}_n) = 0$. Therefore, if $\mathbf{U}_n \in \mathbf{A}_n$, then $d(\mathbf{U}, \mathbf{A}_n) \to 0$, i.e. $\mathbf{U} \in \mathbf{A}_n$ asymptotically with probability one. Hence, $P^*(\mathbf{U} \in \mathbf{A}, \mathbf{U} \notin \mathbf{A}_n, \mathbf{U}_n \in \mathbf{A}_n) \to 0$. Similarly, $P^*(\mathbf{U} \in \mathbf{A}, \mathbf{U} \notin \mathbf{A}_n, \mathbf{U}_n \notin \mathbf{A}_n) \leq P^*(\mathbf{S} \in \mathbf{B}, \mathbf{S}_n \notin \mathbf{B})$. By the triangle inequality,

$$d(\mathbf{S}_n, \mathbf{B}) \leq d(\mathbf{S}_n, \mathbf{S}) + d(\mathbf{S}, \mathbf{B}).$$

By Theorem 6.2 in [Faugeras and Rüschendorf, 2017], $d(\mathbf{S}, \mathbf{S}_n) \to 0$ and $\mathbf{S} \in \mathbf{B}$ implies $\mathbf{S}_n \in \mathbf{B}$ asymptotically with probability one. Hence, $P^*(\mathbf{S} \in \mathbf{B}, \mathbf{S}_n \notin \mathbf{B}) \to 0$. The treatment of (II) is similar. All statements occur w.r.t. the original P-probability one.

2. (a) By definition of the Markov morphisms, one has P^* a.s.

$$\begin{aligned} \mathbf{S} \in \mathbf{B} &\Leftrightarrow \mathbf{R}_C(\mathbf{U}) \in \mathbf{B} \Leftrightarrow \mathbf{U} \in \mathbf{R}_C^{-1}(\mathbf{B}) = \mathbf{Q}_C(\mathbf{B}) = \mathbf{A} \\ &\Rightarrow \mathbf{G}^{-1}(\mathbf{U}) \in \mathbf{G}^{-1}(\mathbf{A}) = \mathbf{Z} \Leftrightarrow \mathbf{X} \in \mathbf{Z} \end{aligned} \tag{25}$$

where $\mathbf{X} \sim F$ also sits on the probability space $(\Omega^*, \mathcal{A}^*, P^*)$ of Theorem 6.2 in [Faugeras and Rüschendorf, 2017], and (25) follows from

the fact that $\mathbf{G}^{-1}(\mathbf{U}) = \mathbf{X}$, P^*-a.s., by definition of the distributional transform. Therefore, $\tau = P(\mathbf{S} \in \mathbf{B}) \leq P(\mathbf{X} \in \mathbf{Z})$. The proof for $\mathbf{Z}_n$ is similar: for $\mathbf{S}_n = \mathbf{R}_{C_n}(\mathbf{U}_n)$,

$$\mathbf{S}_n \in \mathbf{B} \Leftrightarrow \mathbf{U}_n \in \mathbf{A}_n \Rightarrow \mathbf{X}_n^* \in \mathbf{Z}_n \tag{26}$$

The proof that $P^{\mathbf{X}}(z) = P^{\mathbf{X}_n^*}(z_n) = \tau$ follows as in Theorem 2.1 1. (a) from the definitions of the Markov morphisms.
In case $P^{\mathbf{X}}$ is continuous, the distributional transform $\mathbf{X} \to \mathbf{G}(\mathbf{X}, \mathbf{V})$ reduces to $\mathbf{X} \to \mathbf{G}(\mathbf{X})$ and the implications in (25) become equivalences P^* a.s.,

$$\mathbf{S} \in \mathbf{B} \Leftrightarrow \mathbf{U} \in \mathbf{A} \Leftrightarrow \mathbf{X} \in \mathbf{Z}, \tag{27}$$

which yields $P^{\mathbf{X}}(\mathbf{Z}) = \tau$.

(b) One has

$$\begin{aligned} P^{\mathbf{X}}(|\mathbb{1}_{\mathbf{Z}_n} - \mathbb{1}_{\mathbf{Z}}|) &= P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n) + P^*(\mathbf{X} \notin \mathbf{Z}, \mathbf{X} \in \mathbf{Z}_n) \\ &:= (I) + (II) \end{aligned}$$

with

$$\begin{aligned} (I) &= P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n, \mathbf{U}_n \in \mathbf{A}_n) + P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n, \mathbf{U}_n \notin \mathbf{A}_n) \\ &:= (Ia) + (Ib) \end{aligned}$$

By the implication in (26), one has

$$\begin{aligned} (Ia) &:= P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n, \mathbf{U}_n \in \mathbf{A}_n) \leq P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n, \mathbf{X}_n^* \in \mathbf{Z}_n) \\ &\leq P^*(\mathbf{X} \notin \mathbf{Z}_n, \mathbf{X}_n^* \in \mathbf{Z}_n) \end{aligned}$$

The triangle inequality gives

$$d(\mathbf{X}, \mathbf{Z}_n) \leq d(\mathbf{X}, \mathbf{X}_n^*) + d(\mathbf{X}_n^*, \mathbf{Z}_n)$$

By Theorem 6.2 in [Faugeras and Rüschendorf, 2017], $d(\mathbf{X}, \mathbf{X}_n^*) \to 0$ as $n \to \infty$, and $\mathbf{X}_n^* \in \mathbf{Z}_n$ implies $d(\mathbf{X}^*, \mathbf{Z}_n) \to 0$. Therefore, $(Ia) \to 0$.
For $P^{\mathbf{X}}$ continuous, by the equivalences in (26) and (27),

$$\begin{aligned} (Ib) &:= P^*(\mathbf{X} \in \mathbf{Z}, \mathbf{X} \notin \mathbf{Z}_n, \mathbf{U}_n \notin \mathbf{A}_n) \\ &\leq P^*(\mathbf{S} \in \mathbf{B}, \mathbf{S}_n \notin \mathbf{B}) \end{aligned}$$

The triangle inequality gives

$$d(\mathbf{S}_n, \mathbf{B}) \leq d(\mathbf{S}_n, \mathbf{S}) + d(\mathbf{S}, \mathbf{B})$$

By Theorem 6.2 in [Faugeras and Rüschendorf, 2017], $d(\mathbf{S}_n, \mathbf{S}) \to 0$ as $n \to \infty$, and $\mathbf{S} \in \mathbf{B}$ implies $d(\mathbf{S}_n, \mathbf{B}) \to 0$. Therefore, $(Ib) \to 0$.
The treatment of (II) is similar. $P^{\mathbf{X}}(\mathbf{Z}_n) \to \tau$ follows.

(c) See Corollary 3.2.

□

6.2.2 Proof of Theorem 3.1

Proof. One has, with P and P^*-probability one,

$$\begin{aligned} \{\mathbf{X} \in \mathbf{Z_V}\} &= \{\mathbf{G}(\mathbf{X},\mathbf{V}) \in \mathbf{A}\} = \{\mathbf{U} \in \mathbf{A}\} = \{\mathbf{S} \in \mathbf{B}\} \\ \{\mathbf{X} \in \mathbf{Z_{V,}}_n\} &= \{\mathbf{G_n}(\mathbf{X},\mathbf{V}) \in \mathbf{A}_n\} = \{\mathbf{R}_{C_n} \circ \mathbf{G_n}(\mathbf{X},\mathbf{V}) \in \mathbf{B}\}. \end{aligned}$$

Therefore,

$$\begin{aligned} P^*(\mathbf{X} \in \mathbf{Z_V} \setminus \mathbf{Z_{V,}}_n) &= P^{(\mathbf{X},\mathbf{V})}(\{(\mathbf{x},\mathbf{v}) : \mathbf{G}(\mathbf{x},\mathbf{v}) \in \mathbf{A}, \mathbf{G_n}(\mathbf{x},\mathbf{v}) \notin \mathbf{A}_n\}) \\ &= P^*(\mathbf{G}(\mathbf{X},\mathbf{V}) \in \mathbf{A}, \mathbf{G_n}(\mathbf{X},\mathbf{V}) \notin \mathbf{A}_n) \\ &= P^*(\{\mathbf{S} \in \mathbf{B}\} \setminus \{\mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X},\mathbf{V}) \in \mathbf{B}\}) \end{aligned}$$

By ergodicity, for all $\mathbf{x} \in \mathbb{R}^d$, with P-probability one,

$$\begin{aligned} |\mathbf{G}_n(\mathbf{x}-) - \mathbf{G}(\mathbf{x}-)| &\to 0, \\ |\mathbf{G}_n(\mathbf{x}) - \mathbf{G}(\mathbf{x})| &\to 0. \end{aligned}$$

Therefore, for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{v} \in [0,1]^d$, with P-probability one, $\mathbf{G}_n(\mathbf{x},\mathbf{v}) \to \mathbf{G}(\mathbf{x},\mathbf{v})$, which yields

$$\mathbf{G}_n(\mathbf{X},\mathbf{V}) \to \mathbf{G}(\mathbf{X},\mathbf{V}) = \mathbf{U}, \tag{28}$$

with P and P^*-probability one.

For $P^{\mathbf{X}}$ discrete or absolutely continuous, one has that $P^{\mathbf{U}}$ is absolutely continuous w.r.t. λ^d, (see Lemma 6.1 in Appendix 6.1 for the case $P^{\mathbf{X}} \ll \lambda^d$). Therefore, the probability that $\mathbf{U}$ lies on the frontier of its support is null:

$$P^*(\mathbf{U} \in \partial(supp(P^{\mathbf{U}}))) = 0.$$

Together with (28), this imply that there exists some $\epsilon > 0$ and some N_0, s.t. for all $n \geq N_0$, $\mathbf{G}_n(\mathbf{X},\mathbf{V}) \in \mathbf{L} := \mathbf{B}(\mathbf{U},\epsilon) \subset int(supp(P^{\mathbf{U}}))$, with P^* probability one, where $\mathbf{B}(\mathbf{u},\epsilon)$ is the open convex ball of center $\mathbf{u}$ and radius ϵ, together with some $N_1 > N_0$, such that

$$\forall n \geq N_1, \quad \mathbf{G}_n(\mathbf{X},\mathbf{V}) \in \mathbf{K} := \overline{\mathbf{B}(\mathbf{U},\epsilon/2)}, \tag{29}$$

with P^* probability one, where $\overline{\mathbf{B}(\mathbf{u},\epsilon/2)}$ is the closed convex ball of center $\mathbf{u}$ and radius $\epsilon/2$.

Consider the restrictions $P^{\mathbf{U}}|_{\mathbf{L}}$ and $P^{\mathbf{S}}|_{\mathbf{R}_C(\mathbf{L})}$, normalized to be of unit mass, i.e. $P^{\mathbf{U}}|_{\mathbf{L}}(\mathbf{A}) := \frac{P^{\mathbf{U}}(\mathbb{1}_{\mathbf{A}\cap\mathbf{L}})}{P^{\mathbf{U}}(\mathbf{L})}$, for a Borel set $\mathbf{A} \subset [0,1]^d$, and similarly for $P^{\mathbf{S}}|_{\mathbf{R}_C(\mathbf{L})}$. The density of $P^{\mathbf{U}}|_{\mathbf{L}}$ is bounded away from 0 and ∞ (see Remark 1). By Theorem 1.3 in [De Philippis and Figalli, 2014] (if $\mathbf{0} \notin \mathbf{R}_C(\mathbf{L})$)), or Theorem 1.1 in [Figalli, 2018] (if $\mathbf{0} \in \mathbf{R}_C(\mathbf{L})$), there exists two relatively closed sets $\mathbf{\Sigma}_1 \subset \mathbf{R}_C(\mathbf{L})$ and $\mathbf{\Sigma}_2 \subset \mathbf{L}$, of zero Lebesgue measure, s.t. the optimal transport map $\mathbf{T} : \mathbf{R}_C(\mathbf{L}) \setminus \mathbf{\Sigma_1} \to \mathbf{L} \setminus \mathbf{\Sigma_2}$ sending $P^{\mathbf{S}}|_{\mathbf{R}_C(\mathbf{L})}$ towards $P^{\mathbf{U}}|_{\mathbf{L}}$ is a bi-Hölder homeomorphism. Obviously, the optimal transport map of

the restricted measures is the restriction of the optimal transport map of the unrestricted measures, i.e. $\mathbf{T} = \mathbf{Q}_C|_{\mathbf{R}_C(\mathbf{L})}$, see e.g. Proposition 2.3 in [Cuesta-Albertos et al., 1997]. One has therefore that its inverse $\mathbf{R}_C$ is continuous on $\mathbf{L}$, up to a null set. A fortiori, $\mathbf{R}_C$ is continuous on $\mathbf{K}$, up to a null set. Moreover, also by restriction, condition (C) of Theorem A.2 in [Chernozhukov et al., 2017] is satisfied between $P^{\mathbf{U}}|_{\mathbf{K}}$ and $P^{\mathbf{S}}|_{\mathbf{R}_C(\mathbf{K})}$, which yields uniform convergence on $\mathbf{K}$ of $\mathbf{R}_{C_n}$ towards $\mathbf{R}_C$, up to a null set.

These two statements entails the continuous convergence of $\mathbf{R}_{C_n}$ towards $\mathbf{R}_C$ on $\mathbf{K}$, up to a null set. By (28), (29), and the extended continuous mapping, Theorem 1.11.1 in [van der Vaart and Wellner, 1996], one has that, with $P-$ probability one,

$$\mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}, \mathbf{V}) \overset{P^* a.s.}{\longrightarrow} \mathbf{R}_C \circ \mathbf{G}(\mathbf{X}, \mathbf{V}) = \mathbf{S}.$$

This implies that

$$P^*(\mathbf{X} \in \mathbf{Z}_{\mathbf{V}} \setminus \mathbf{Z}_{\mathbf{V},n}) = P^*(\{\mathbf{S} \in \mathbf{B}\} \setminus \{\mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}, \mathbf{V}) \in \mathbf{B}\}) \to 0,$$

since $\mathbf{B}$ is a continuity set of $P^{\mathbf{S}}$. One has similarly, for the other inclusion,

$$P^*(\mathbf{X} \in \mathbf{Z}_{\mathbf{V},n} \setminus \mathbf{Z}_{\mathbf{V}}) \to 0,$$

which yields the result. □

6.2.3 Proof of Proposition 4.1

Proof. The proof follows the main arguments of Theorem 6.2 in [Faugeras and Rüschendorf, 2017], with minor modifications. On the auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$ on which (10) holds, $\mathbf{h}_n \downarrow \mathbf{0}$ and (10) in (21) yields, with P-probability one,

$$\hat{\mathbf{X}}_n^* \overset{P^* a.s.}{\longrightarrow} \mathbf{X}. \tag{30}$$

We have the decomposition,

$$\begin{aligned} \hat{\mathbf{U}}_n - \mathbf{U} &= \hat{\mathbf{G}}_n(\hat{\mathbf{X}}_n^*) - \mathbf{G}(\hat{\mathbf{X}}_n^*) + \mathbf{G}(\hat{\mathbf{X}}_n^*) - \mathbf{G}(\mathbf{X}) \\ &\leq ||\hat{\mathbf{G}}_n - \mathbf{G}||_\infty + \mathbf{G}(\hat{\mathbf{X}}_n^*) - \mathbf{G}(\mathbf{X}), \end{aligned} \tag{31}$$

where the operations are to be understood componentwise. By ergodicity, one has that $\forall \mathbf{x} \in \mathbb{R}^d$, $\mathbf{G}_n(\mathbf{x}) \to \mathbf{G}(\mathbf{x})$. Since $\mathbf{G}$ is continuous, this implies by Polya's Theorem [Pólya, 1920], a Glivenko-Cantelli type Theorem for $\mathbf{G}_n$, i.e.

$$||\mathbf{G}_n - \mathbf{G}||_\infty \to 0,$$

with P-probability one. Obviously $P^{\mathbf{h}_n \mathbf{W}} \overset{d}{\to} \delta_{\mathbf{0}}$, so one has, as in [Winter, 1973] Theorem 2, that

$$||\hat{\mathbf{G}}_n - \mathbf{G}||_\infty \to 0, \tag{32}$$

with P-probability one. (See also [Yamato, 1972/73], [Singh et al., 1983], [Yukich, 1989] where (32) was obtained in the i.i.d. case, but extends readily in the

ergodic setup. See [Rao, 1962]). Hence, (32) in (31), and continuity of $\mathbf{G}$ with (30) entails

$$\hat{\mathbf{U}}_n \overset{P^* a.s.}{\longrightarrow} \mathbf{U},$$

with P-probability one. The rest of the proof is as Theorem 6.2 in [Faugeras and Rüschendorf, 2017]: by [Cuesta-Albertos et al., 1997] Theorem 3.4, one gets

$$\hat{\mathbf{S}}_n \overset{P^* a.s.}{\longrightarrow} \mathbf{S},$$

with P-probability one. □

7 Appendix 2: Technical details for the simulation methods

7.1 Implementation with Linear Programming

We implemented the linear programming version of the algorithm using the LinearProgramming command of Mathematica, which yields the optimal transportation matrix. The optimal transportation map is then obtained by picking the position of the non-zero values of the transportation matrix. Figure 15 shows an implementation of the algorithm for a sample of $n = 800$ points from a standard bivariate Gaussian distribution. The (red circles) points of the discretized unit disk are sent by the optimal transportation map to a (unique) realization of the empirical copula (blue filled squares). We drew arrows from the $\tau = 0.5$ quantile contour of the unit disk towards the copula space.

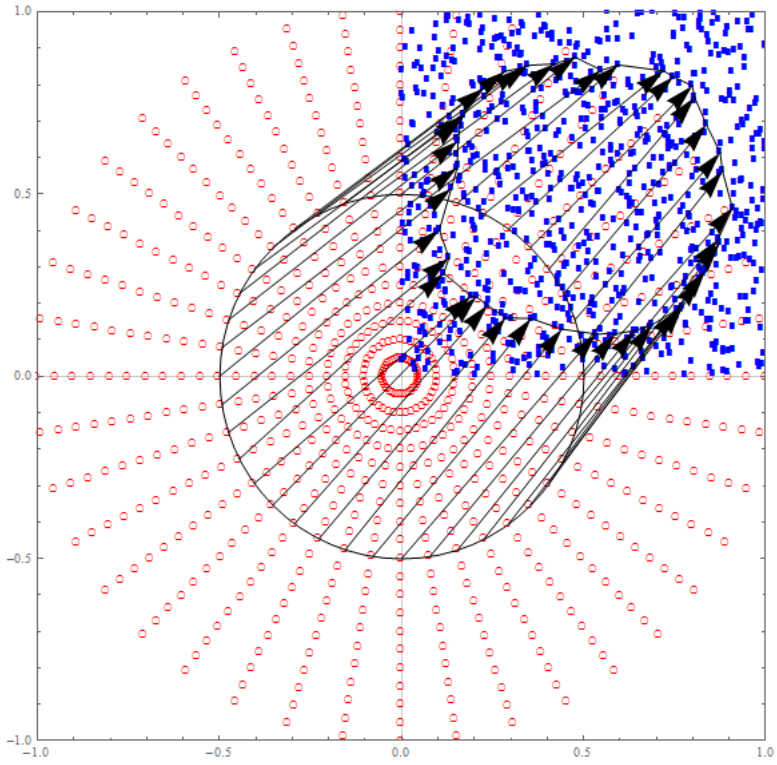

Figure 15: Optimal Transportation from the discretized unit disk (Red circles) to the empirical copula points (Blue points). The arrows show the optimal transportation map sending the red points on the circle of radius $\tau = 0.5$ towards their counterparts in the (empirical) copula space. Computed from $n = 800$ i.i.d. observations from a bivariate standard Gaussian distribution.

Figure 16 complements Figure 15 by displaying the resulting empirical quantile area $\mathbf{A}_n$ in the copula space as represented by the (red filled squares) points. We naively joined the copula points on the $\tau = 0.5$ quantile contour with a straight line to obtain a continuous quantile contour outside of the sample copula points. A more sophisticated approach would be to use the cyclically monotone interpolant proposed in [del Barrio et al., 2018].

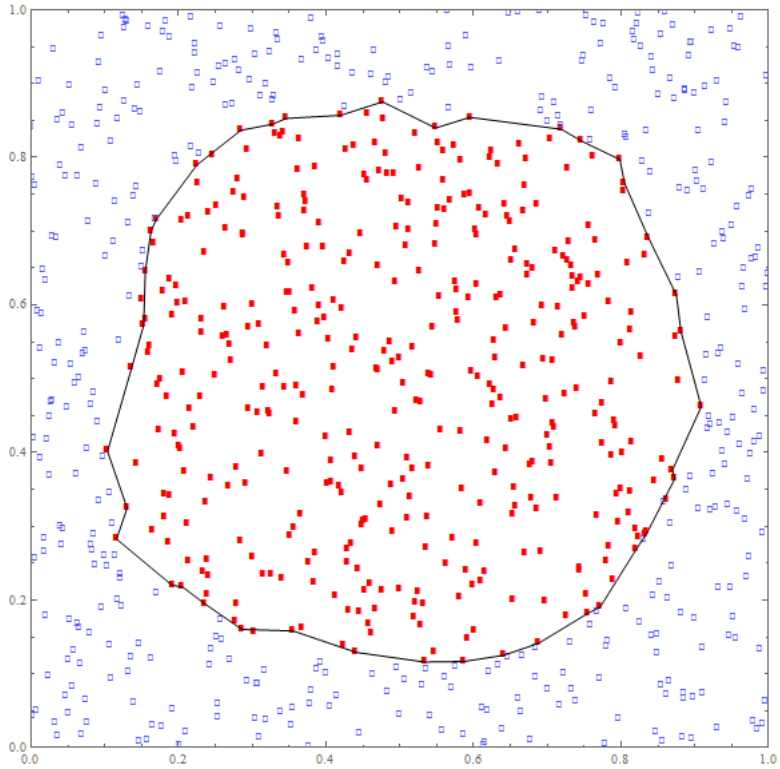

Figure 16: Empirical quantile area $\mathbf{A}_n$ of level $\tau = 0.5$ in the copula space (red filled squares). The blue empty squares points are the remaining empirical copula sample points.

These (red filled square) points in the copula space $[0,1]^2$ are then mapped by the marginal quantile transform map $\mathbf{G}_n^{-1}$ towards the original sample points from $P^{\mathbf{X}}$. One then obtains in Figure 17 the empirical $\tau = 0.5$ quantile area $\mathbf{Z}_n$ in the original sample space $\mathbb{R}^2$, represented as red filled circles. The theoretical quantile disk of level $\tau = 0.5$ is represented by the shaded disk, and the remaining sample points outside the central quantile region are displayed with blue circles.

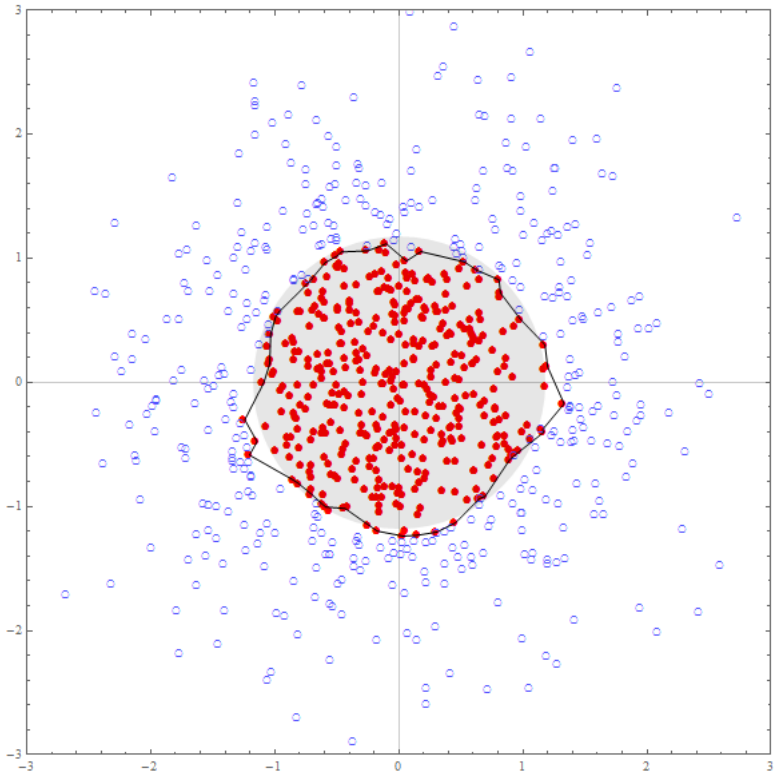

Figure 17: Empirical quantile area $\mathbf{Z}_n$ of asymptotic level $\tau = 0.5$ in the sample space (red filled circles inside the broken line). The theoretical quantile disk of level $\tau = 0.5$ is represented by the shaded disk.

7.2 Implementation with the Entropic regularization method and Sinkhorn algorithm

The linear programming code slows down as the sample size increase and definitely runs into memory trouble for sample size n exceeding 800 points on our computer. In addition, it is known that the best algorithms for optimal assignments have at least a $O(n^3)$ computational cost, hence all known algorithms that can solve the optimal transport problem scale at least super-cubicly in n. To alleviate these issues, [Cuturi, 2013] recently proposed to use an entropic regularization of the optimal transport cost, which we briefly explain. See Peyré et al. [2019] for details.

For two probability measures $P^{\mathbf{X}}$ and $P^{\mathbf{Y}}$, the entropic regularisation method solves

$$W_\gamma(P^{\mathbf{X}}, P^{\mathbf{Y}}) := \inf_{P^{(\mathbf{X},\mathbf{Y})} \in \Pi(P^{\mathbf{X}}, P^{\mathbf{Y}})} \int ||\mathbf{x} - \mathbf{y}||^{\mathbf{2}} dP^{(\mathbf{X},\mathbf{Y})}(\mathbf{x}, \mathbf{y}) + \gamma H(P^{(\mathbf{X},\mathbf{Y})}), \tag{33}$$

where the infimum is on the set $\Pi(P^{\mathbf{X}}, P^{\mathbf{Y}})$ of all joint distributions $P^{(\mathbf{X},\mathbf{Y})}$ with marginals $P^{\mathbf{X}}$ and $P^{\mathbf{Y}}$. $\gamma > 0$ is a parameter and $H(P^{(\mathbf{X},\mathbf{Y})})$ is the entropy of the coupling distribution (in practice, $P^{(\mathbf{X},\mathbf{Y})}$ is a discrete distribution so the entropy is well-defined). Heuristically, by adding a penalization term which is minimal for "spread-out" distributions, the entropic regularization method forces the transportation plan of (33) to be induced by a non-degenerate Markov morphism instead of a mapping, in a fashion reminiscent of Kantorovich's relaxation of Monge Problem. One can show, see e.g. [Peyré et al., 2019] Chapter 4, that the regularized problem is strongly convex (hence stabilizing computations), and that its solution converges exponentially to the original, nonregularised (i.e. for $\gamma = 0$) optimal transportation solution, as $\gamma \to 0$. Moreover, (33) can be rewritten as a Kullback-Leibler projection of the Gibbs distribution $\zeta(d\mathbf{x}, d\mathbf{y}) := e^{\frac{-||\mathbf{x}-\mathbf{y}||^2}{\gamma}} d\mathbf{x}d\mathbf{y}$,

$$W_\gamma(P^{\mathbf{X}}, P^{\mathbf{Y}}) = \inf_{P^{(\mathbf{X},\mathbf{Y})} \in \Pi(P^{\mathbf{X}}, P^{\mathbf{Y}})} KL(P^{(\mathbf{X},\mathbf{Y})}, \zeta). \tag{34}$$

The solution of (34) is a diagonal scaling of ζ, which can be found efficiently through the well-known IPFP [Deming and Stephan, 1940]/ Sinkhorn [Sinkhorn, 1964, 1967, Sinkhorn and Knopp, 1967] algorithm, see [Peyré et al., 2019] Chapter 4. For discrete $P^{\mathbf{X}}$ and $P^{\mathbf{Y}}$, the optimal $P^{(\mathbf{X},\mathbf{Y})}$ is akin to a matrix which can be simply computed by vector-matrix and elementary operations. In practice, the regularisation parameter γ must be small enough so that the solution be close to the unregularised solution sought, but large enough to prevent division by zero. We used $\gamma = 0.01$ in all our simulations.

From the optimal transportation matrix solution of (33), one can obtain the optimal transportation map in several ways:

- **mode method:** Since the solution of (33) converges, as $\gamma \to 0$, to the optimal transportation plan, which is, for discretized distributions with the same number of points, a (rescaled) permutation matrix, one can get the optimal transportation map by selecting the position $\mathbf{y}$ corresponding to the highest value on each line $\mathbf{X} = \mathbf{x}$, i.e. by taking the mode of $P^{\mathbf{Y}|\mathbf{X}=x}$.

- **barycenter method:** [Peyré et al., 2019] recommends to extract an (approximate) optimal transportation map by computing the conditional expectation $E[\mathbf{Y}|\mathbf{X} = \mathbf{x}]$. This amounts to computing a barycenter of the $\mathbf{Y}$ points, weighted by the regularized optimal transportation solution of (33). This creates an additional smoothing/convexification effect, which somehow makes the quantile contours more regular and visually appealing. However, it only gives an approximate transportation map, since the cloud of barycenters points obtained is slightly different from the original cloud of destination points, see Figure 19 below. Also, for nonconvex quantile regions, the convexification induced by taking barycenters is undesirable, as it modifies its shape.

The following figures illustrate the procedure, with the same experimental setting as in Figures 15, 16, 17. The number of iterations in Sinkhorn's algorithm is controlled by computing the relative error of the marginals of $P^{(\mathbf{X},\mathbf{Y})}$ to their target marginal distribution. Figure 18 shows the relative error in the matching of the marginals of the transportation plan as the number of iteration increases: we found that 300 iterations were enough to guarantee an error around 10^{-12}, in all our simulations.

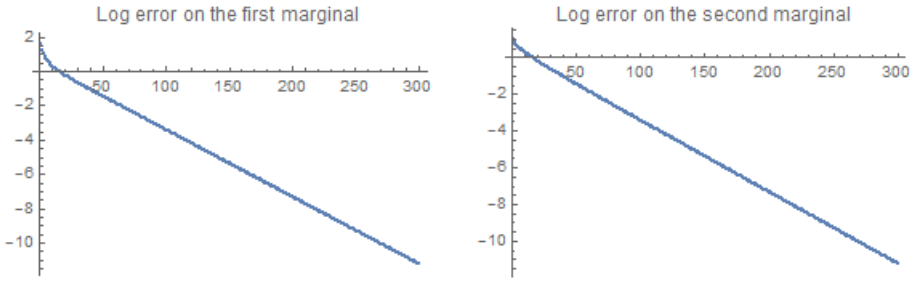

Figure 18: Log error in Sinkorn's algorithm on the matching of the first marginal of the optimal transportation plan (left panel) and the second marginal (right panel). $n = 800$.

Figure 19 illustrates how the optimal quantile areas are computed: the left panel shows the discretized unit disk with its $\tau = 0.5$ spherical depth area. These points are then mapped into the copula by the optimal approximate transportation map derived from Sinkhorn algorithm + barycenter method (center panel), which are subsequently mapped by the empirical marginal quantile transform $\mathbf{G}_n^{-1}$ (right panel).

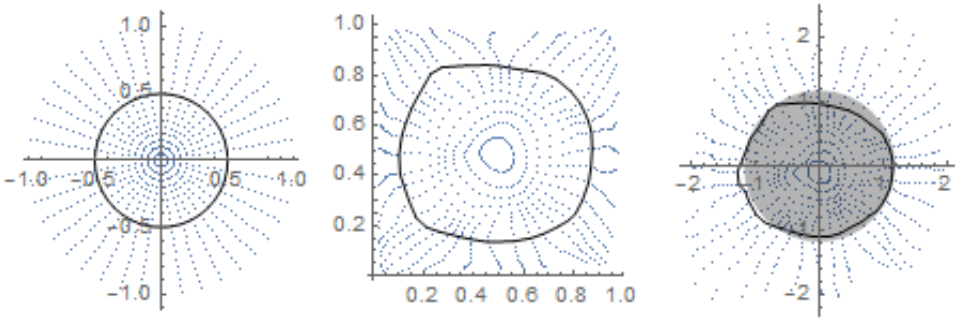

Figure 19: Discretized unit disk with $\tau = 0.5$ depth area (left panel). Mapping of the unit disk in the copula space (center panel) and in the sample space (right panel). $n = 800$.

Figure 20 illustrates the $\tau = 0.5$ quantile areas obtained: the left panel show

the $\tau = 0.5$ quantile area in the copula space together with the empirical copula points, while the right panel shows the $\tau = 0.5$ quantile area in the sample space together with the theoretical quantile area (shaded disk) and the original data points.

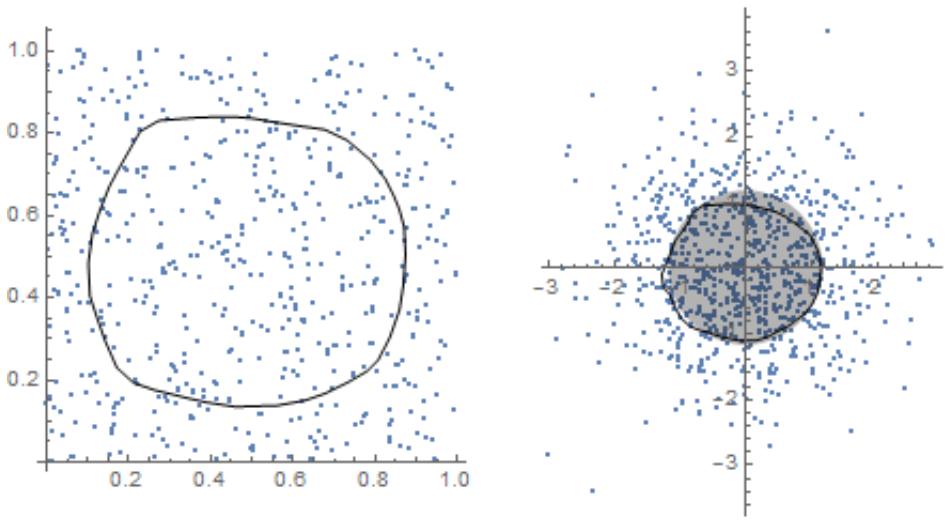

Figure 20: $\tau = 0.5$ quantile area in the copula space with the sample from the empirical copula (left panel), $\tau = 0.5$ quantile area in the sample space with the original sample data and the theoretical quantile area (shaded disk)(right panel). $n = 800$.

Note the slight discrepancies between the original sample points (in the observational space and in the copula space) in Figure 20 and the reconstructed points obtained by the barycenter method, as in Figure 19.

For the smoothed approach of Section 4, we show in Figure 21 its implementation using the same Sinkorn algorithm + barycenter method as in Figure 20. We get a $\tau = 0.5$ empirical quantile area very close to the theoretical one (represented as a the shaded disk).

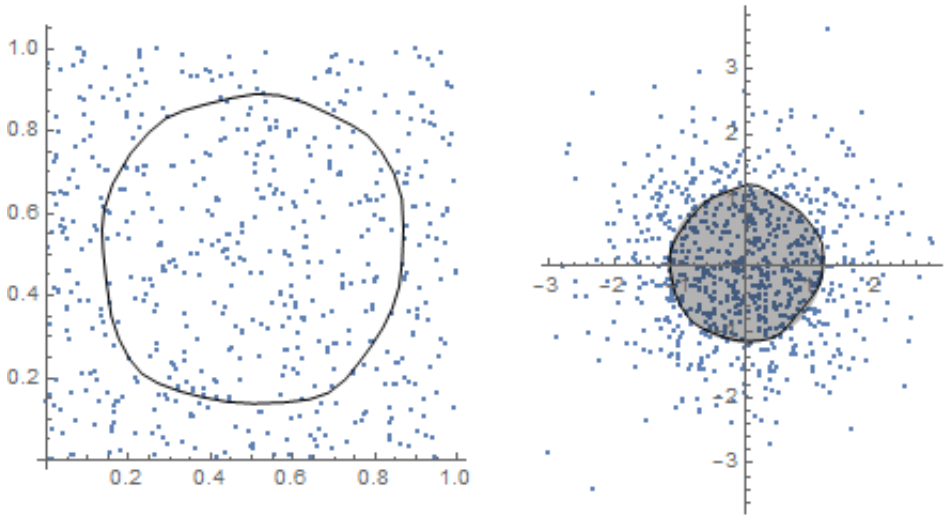

Figure 21: $\tau = 0.5$ smoothed quantile area in the copula space with the sample from the smoothed empirical copula (left panel), $\tau = 0.5$ smoothed quantile area in the sample space with the original sample data and the theoretical quantile area (shaded disk)(right panel). $n = 800$. Implementation= Sinkhorn + barycenter method

Eventually, for comparison purposes with the right panel of Figure 21, we show in Figure 22 the implementation of the smoothed approach of of Section 4 with the Sinkorn algorithm + mode method.

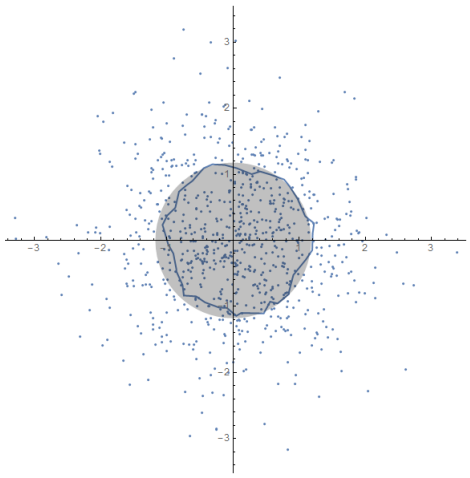

Figure 22: $\tau = 0.5$ smoothed quantile area in the sample space with the original sample data and the theoretical quantile area (shaded disk). $n = 800$. Implementation= Sinkhorn + barycenter method

7.3 Comparison of the algorithms

Comparing Figures 16 and 17 (randomized depth approach with Linear programming), with Figure 20 (randomized depth approach with Sinkhorn algorithm + barycenter method), and with Figure 21 (smoothed depth approach with Sinkhorn algorithm + barycenter method), the randomized depth approach of Sections 3 and smoothed depth approach of Section 4 give close results, with the smoothed depth approach yielding the most visually appealing results and closest to the theoretical quantile area.

Comparing the implementations of the mass transportation step, the entropy regularised version with the Sinkhorn algorithm is the most efficient (e. g. computations take about 35 seconds for 5000 points, and a couple of seconds for 800 points) and is highly recommended compared to the linear programming method (which breaks down for more than 800 points). For computing the optimal transportation map, the barycenter method works well for distributions with "nice", convex, connected support, while the mode method, by mapping to an exact point of the sample, thus giving a depth area falling inside the support and respecting its geometry, should be preferred.

On the computational side, the smoothing approach of Section 4 requires more time compared to the randomized depth approach, as one needs to optimize the bandwidths and do more complicated computations.