Functional, randomized and smoothed multivariate quantile regions

Olivier P. Faugeras[b,*], Ludger Rüschendorf[1]

[a] *Université Toulouse 1 Capitole, Toulouse School of Economics, 1 Esplanade de l'université, 31080 Toulouse Cedex 06, France.*
[b] *Albert-Ludwigs University of Freiburg, Eckerstrasse 1, Zimmer 242, D - 79104 Freiburg, Germany.*

Abstract

The mass transportation approach to multivariate quantiles in Chernozhukov et al. [3] was modified in Faugeras and Rüschendorf [8] by a two steps procedure. In the first step, a mass transportation problem from a spherical reference measure to the copula is solved and combined in the second step with a marginal quantile transformation in the sample space. Also, generalized quantiles given by suitable Markov morphisms are introduced there.

In the present paper, this approach is further extended by a functional approach in terms of membership functions, and by the introduction of randomized quantile regions. In addition, in the case of continuous marginals, a smoothed version of the empirical quantile regions is obtained by smoothing the empirical copula. All three extended approaches give empirical quantile ares of exact level and improved stability. The resulting depth areas give a valid representation of the central quantile areas of a multivariate distribution and provide a valuable tool for their analysis.

Keywords: copula, depth area, mass transportation, vector quantiles.
2020 MSC: Primary 62G15, Secondary 62H05

1. Introduction

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

In Chernozhukov et al. [3], a mass transportation approach has been proposed to the definition of multivariate quantiles and depth areas. The basic idea 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 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 McCann [12]. This was used in Hallin et al. [11] to defining a multivariate quantile function at the observed sample points by solving an empirical optimal matching problem. (See also Ghosal and Sen [10]). 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 [5], 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 [8], the mass transportation approach of Chernozhukov et al. [3] was generalized by defining the "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 was 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. [3], and Faugeras and Rüschendorf [8] are as follows: for a random vector $X \in \mathbb{R}^d$ with c.d.f. F, we denote by $G = (G_1, \ldots, G_d)$ the vector of marginal cdfs,

*Corresponding author. Email address: `olivier.faugeras@tse-fr.eu`

i.e., $G_i(x_i) := F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$. We consider the spherical uniform distribution P^S of a r.v. S on the unit ball $B_1 = \{s \in \mathbb{R}^d; \|s\| \leq 1\}$. Identifying $S \in B_1$ with the pair (r, a), where the radial part $r \sim U_{[0,1]}$ is uniformly distributed on the interval $[0, 1]$, and the angular part a is uniformly distributed on the unit sphere in $\mathbb{R}^d$, such a spherical distribution P^S gives natural depth regions of P^S mass τ as the balls B_τ of radius τ, since $\mathrm{P}^S(B_\tau) = \tau$. The basic idea of Chernozhukov et al. [3] and Hallin et al. [11] is then to transform these balls B_τ into depth regions of level τ in the observational space, by setting

$$A_\tau := Q_F(B_\tau), \tag{1}$$

where Q_F is the optimal mass transportation mapping obtained by pushing forward P^S to P^X by optimal transport, i.e.,

$$Q_F \# \mathrm{P}^S = \mathrm{P}^X. \tag{2}$$

Q_F is the optimal mapping induced by the solution of the Monge-Kantorovich Wasserstein metric,

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

where the infimum is over all couplings γ of $(\mathrm{P}^S, \mathrm{P}^X)$. (In Hallin et al. [11], Q_F is defined as the gradient of the convex potential in the qualitative McCann [12] approach, so second moment assumptions are not needed for the existence of Q_F). By noting R_F the corresponding "rank" reciprocal optimal mass transportation mapping, i.e.,

$$R_F \# \mathrm{P}^X = \mathrm{P}^S, \tag{3}$$

one gets that

$$\mathrm{P}^X(A_\tau) = \mathrm{P}(Q_F(S) \in Q_F(B_\tau)) = \mathrm{P}(S \in R_F \circ Q_F(B_\tau)) = \mathrm{P}^S(B_\tau) = \tau. \tag{4}$$

For (2), (3) and thus (4) to hold, one needs regularity conditions on the distribution P^X (see Rüschendorf and Rachev [16], Brenier [2], and McCann [12]). These regularity assumptions are, e.g., not satisfied when P^X has a discrete component, so, in particular, when P^X is replaced by the empirical measure $\mathrm{P}^{X_n^*} := n^{-1} \sum_{i=1}^n \delta_{X_i}$ based on an ergodic sample realization $X_1, X_2, \ldots$ of X. Thus, the empirical versions of the depth areas in Chernozhukov et al. [3] require a smoothed version of the empirical measure and regularity assumptions. This kind of regularity assumptions is not needed anymore by the approach of Faugeras and Rüschendorf [8], who introduce a copula step and do instead the optimal transportation between the reference spherical measure P^S and a/the copula measure P^{U} associated to P^{X}. Thus, this step is based on an analysis of the dependence structure. In a second step, the copula measure P^U is transformed to P^X by the marginal quantile functions, resp. in opposite direction by the distributional transform. This combined copula-mass transportation approach modifies (1) and is summarized as follows:

1. Transform $X \sim F$ into a copula representer $U = G(X, V)$, whose c.d.f. C is a copula of F. Here, V is a vector with i.i.d. components uniformly distributed on $[0, 1]$, and

$$X \mapsto G(X, V) = (G_1(X_1, V_1), \ldots, G_d(X_d, V_d)) \tag{5}$$

   is the multivariate marginal distributional transform: $G_i(x_i, v_i) = P(X_i < x_i) + v_i P(X_i = x_i)$. If P^X is discrete, this transformation randomizes the jumps of each components X_i so that each U_i is uniformly distributed on $[0, 1]$. In this way, it associates to P^X a particular copula P^U. If P^X is continuous, it reduces to the probability integral transform $G(X)$ and the copula is unique, see Faugeras and Rüschendorf [8]. In both cases, under the assumptions of Lemma 9, P^U is absolutely continuous so that the optimal transportation plans between P^U and P^S are induced by mappings.
2. Transport P^U into the spherical reference distribution P^S via optimal mass transportation maps Q_C, R_C, i.e.,

$$Q_C \# \mathrm{P}^S = \mathrm{P}^U, \quad R_C \# \mathrm{P}^U = \mathrm{P}^S. \tag{6}$$

3. The balls B_τ of P^S-mass τ, are mapped into depths regions A_τ at the copula level of P^U-mass τ:

$$A_\tau := Q_C(B_\tau), \quad \mathrm{P}^U(A_\tau) = \tau. \tag{7}$$

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

$$Z_\tau := G^{-1}(A_\tau). \tag{8}$$

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^X level, i.e.,

$$\mathrm{P}^X(Z_\tau) = \tau.$$

The procedure can be formulated in terms of transformations of probability measures via Markov kernels, as in Fig. 1. We denote by $\mathcal{Q}_C, \mathcal{R}_C, \mathcal{G}^{-1}$ the degenerate Markov morphisms induced by the mappings Q_C, R_C, G^{-1}, e.g., $\mathcal{Q}_C(S, .) = \delta_{Q_C(S)}(.)$. $\mathcal{D}$ denotes the (non-degenerate) Markov kernel corresponding to the distributional transform $X \to U = G(X, V)$, i.e., $D(x, .) = \mathrm{P}^{U|X=x}(.) = \mathrm{P}^{G(x,V)}(.)$, since V is chosen independent of X.

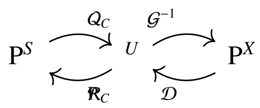

Fig. 1: Markov morphisms of probability measures

Empirical central quantile areas require a modified treatment, as described by the following diagram in Fig. 2, with notations similar as in Fig. 1:

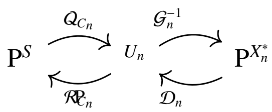

Fig. 2: Markov morphisms of empirical probability measures

Here, we have:

1. $X_n^* \sim F_n$ is, conditionally on the sample, a bootstrap replication, distributed according to the empirical c.d.f. F_n. Set $U_n = G_n(X_n^*, V)$ as an empirical copula representer, whose c.d.f. C_n is a copula of F_n, obtained by the specific empirical distributional transform

$$X_n^* \mapsto G_n(X_n^*, V), \tag{9}$$

with corresponding Markov kernel $\mathcal{D}_n$.
2. Q_{C_n}, R_{C_n} are the optimal transportation maps between the empirical copula and reference measures, viz.

$$Q_{C_n} \# \mathrm{P}^S = \mathrm{P}^{U_n}, \quad R_{C_n} \# \mathrm{P}^{U_n} = \mathrm{P}^S. \tag{10}$$

3. Defining, similarly to (7) and (8) the depth regions

$$A_n := Q_{C_n}(B_\tau), \quad Z_n := G_n^{-1}(A_n), \tag{11}$$

for a fixed level $0 < \tau < 1$, one obtains $\mathrm{P}^{U_n}(A_n) = \tau$, but in general Z_n is not of exact $\mathrm{P}^{X_n^*}$-level τ. If τ is not in the range of values $\{0, 1/n, \ldots, 1\}$ of the empirical measure $\mathrm{P}^{X_n^*}$, it is clear that Z_n can not be a set of $\mathrm{P}^{X_n^*}$-mass τ. A more fundamental reason comes from the fact that since $\mathcal{D}$ (in the case of a discrete X) and $\mathcal{D}_n$ (in every case) are non-degenerate Markov kernels (due to the presence of the randomizer V in the distributional transforms), sets like A (resp. 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.

1.2. Outline

As a consequence of the above given discussion, for a general distribution and also at the empirical level, the natural constructions in (11) lead to depth domains in the observation space of inexact level, which also induces problems in the proof of the consistency result of Corollary 6.4 in Faugeras and Rüschendorf [8]. 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 introduce instead membership functions. These membership functions generalize indicator functions of sets and specify for any point 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^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^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.

The different versions of the depth areas introduced turn out as a reliable and useful data analytical tool for the analysis of multivariate distributions. Section 5 illustrates by numerical simulations some of the remarkable features of the proposed quantile regions, in particular the ability of the proposed methods to capture non-convex or even non-connected domains, and its equivariance w.r.t general monotone transformations of the marginals. A summary of the main features of the proposed approaches is presented in the Conclusion. Proofs of the main results are relegated to Section 6.

1.3. Setting and notation

We interpret operations between vectors componentwise. P^X will stand for the law associated with its representing variable X. We follow the framework and assumptions used in Faugeras and Rüschendorf [8]:

- Ergodicity hypothesis: let $X_1, X_2, \dots$ be an ergodic sample realization of P^X defined on some probability space $(\Omega, \mathcal{A}, P)$. It will be understood that all random variables defined in this article, S, U, X, V, and S_n, U_n, X_n^* of Sections 2-4, $\hat{S}_n, \hat{U}_n, \hat{X}_n^*$ 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 [8] holds, i.e., on the auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$ which allows to construct representers X_n^* of the empirical measure, resp. X of P^X, s.t., with P-probability one, holds
$$X_n^* \stackrel{P^* a.s.}{\longrightarrow} X, \tag{12}$$
see steps one and two in the proof of Theorem 6.2 in Faugeras and Rüschendorf [8]. (Compared to the notation in Faugeras and Rüschendorf [8], we simplify notation and drop the * in the X^*, which reminded that X^* representing P^X was defined on this auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$). We will also denote, for simplicity, e.g., by P^S the law of S, and not by P^{*S}.

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

  In the absolutely continuous case, $P^X \ll \lambda^d$, we will occasionally have to make the following mild regularity assumption on the support of P^X:
$$P^X(\mathring{E}) = 1, \tag{13}$$
  where $E = \{x \in \mathbb{R}^d : f(x) > 0\}$ and $\mathring{E}$ denotes its interior.

2. Depth areas defined by membership functions

As described in Section 1, the mass transformations between probability measures Q_C, $\mathcal{R}_{C_n}$ and the marginal quantile transformations $\mathcal{G}^{-1}$, $\mathcal{D}$, as well as their empirical counterparts (see 1 and 2), are given by Markov kernels and in general are not defined in terms of degenerate kernels, i.e., by measurable functions. Only kernels $\mathcal{K} : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$ arising from a transformation obtained by mass transportation of a non-atomic probability measure $\mathrm{P} \in \mathcal{P}(\mathcal{X})$, can be identified with a measurable mapping $f : \mathcal{X} \to \mathcal{Y}$ as

$$\mathcal{K}(x, dy) = \delta_{f(x)}(dy).$$

As a consequence, a non-degenerate Markov kernel $\mathcal{K}$ transforms a measurable set $A \subset \mathcal{Y}$ into a measurable function $x \to \mathcal{K}(x, A)$, which is generally not binary valued and hence can not be identified with the indicator function of a measurable set in $\mathcal{X}$. Therefore, the use of these kernels to define depth areas leads to a serious regularity condition (see the discussion in Remark 3) for the mass transportation approach to depth sets.

This problem leads us in our first approach to replace depth sets by membership functions. A membership function describes, for a point x in the space considered, the probability with which x belongs to a (randomized) depth area. More precisely, for fixed level $0 < \tau < 1$, and using the simplified notation $B = B_\tau$, $A = A_\tau$, $Z = Z_\tau$, we define the notion of membership function as follows:

Definition 1 (Membership functions)**.** Define the membership functions b, a, z as

(i) $b(s) = \mathbb{1}_B(s)$, for $s \in B_1$ on the spherical reference space,

(ii) $a = \mathcal{R}_C b$ on the copula space,

(iii) $z = \mathcal{D}a = \mathcal{D}\mathcal{R}_C b$ on the observation space.

Similarly, the membership functions for the empirical versions are defined by $a_n = \mathcal{R}_{C_n} b$, on the copula space, $z_n = \mathcal{D}_n a_n = \mathcal{D}_n \mathcal{R}_{C_n} b$, on the observation space.

By definition, one has $\mathrm{P}^S(b) = \int b(s) d\mathrm{P}^S(s) = \tau$ and

$$a(u) = \int \mathcal{R}_C(u, ds) b(s), \quad u \in [0,1]^d, \tag{14}$$

$$z(x) = \int \mathcal{D}(x, du) a(u) = \int \left(\int \mathcal{R}_C(u, ds) b(s) \right) \mathcal{D}(x, du), \quad x \in \mathbb{R}^d,$$

and similarly for the empirical versions. In diagram form, these relations are given in Fig. 3, the dual diagram of Fig. 1.

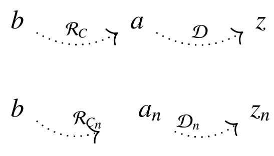

Fig. 3: Membership functions

Note that by Lemma 9, the copula measures P^U, P^{U_n} are absolutely continuous, therefore $\mathcal{R}_C$, $\mathcal{R}_{C_n}$ are degenerate Markov kernels induced by the mappings R_C, R_{C_n}. Hence, the membership functions a, a_n are proper indicator functions, defining proper subsets $A = Q_C(B), A_n = Q_{C_n}(B)$ as in (7) and (11) in the copula space. Indeed, by (14),

$$a(u) \quad = \quad \int_B \delta_{R_C(u)}(ds) = \mathbb{1}_{R_C(u) \in B} = \mathbb{1}_{u \in R_C^{-1}(B)} = \mathbb{1}_{u \in Q_C(B)} = \mathbb{1}_A(u)$$

This is a main feature of the copula and mass transportation approach to depth sets, compared to the direct approach of Chernozhukov et al. [3]. In the observational space, the membership functions z (in the case of non-Lebesgue continuous P^X), and z_n (in any case) can no longer be identified with the deterministic depth set Z and Z_n of (11).

The following theorem is a main result of the paper. It clarifies the coverage probability of membership functions and depth sets, establishes the consistency of their empirical versions towards their population versions, and corrects an inaccuracy in Corollary 6.4 in Faugeras and Rüschendorf [8]:

Theorem 2.

1. *At the copula level:*
   (a) *A is exactly of P^U-mass τ, A_n is exactly of P^{U_n}-mass τ:*

   $$\mathrm{P}^U(A) = \mathrm{P}^{U_n}(A_n) = \tau.$$

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

   $$\mathrm{P}^U(A_n) \to \tau, \quad \text{as } n \to \infty.$$

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

   $$\mathrm{P}^U(|a_n - a|) = \mathrm{P}^U(A_n \triangle A) \to 0, \tag{15}$$

   as $n \to \infty$, with $\mathrm{P}-$probability one.
2. *At the observational X level:*
   (a) *the depth set Z is of P^X-mass at least τ, the depth set Z_n is of $\mathrm{P}^{X_n^*}$-mass at least τ,*

   $$\mathrm{P}^X(Z) \geq \tau, \quad \mathrm{P}^{X_n^*}(Z_n) \geq \tau,$$

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

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

   If P^X is absolutely continuous, then Z is exactly of P^X-mass τ: $\mathrm{P}^X(Z) = \tau$.
   (b) *With $\mathrm{P}-$probability one, if P^X is absolutely continuous, the P^X symmetric distance between Z and Z_n becomes asymptotically negligible,*

   $$\mathrm{P}^X(Z_n \triangle Z) \to 0,$$

   so that Z_n is asymptotically of P^X-mass τ:

   $$\mathrm{P}^X(Z_n) \to \tau.$$

   (c) *With $\mathrm{P}-$probability one, if i) P^X is discrete or ii) $\mathrm{P}^X \ll \lambda^d$ and (13) holds, then the $L_1(\mathrm{P}^X)$ distance between the membership functions z_n and z, is asymptotically null:*

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

   as $n \to \infty$.

3. Randomized depths areas

The interpretation in Section 2 of the Markov kernels $\mathcal{D}$, resp. $\mathcal{D}_n$, by means of the multivariate distributional transforms (5), resp. (9), gives us a tool to define directly a random depth area as a random set.

Definition 3 (Randomized depth area)**.** Let V be the randomizer used in the distributional transforms.

a) Define, for fixed level $0 < \tau < 1$, and for $v \in [0,1]^d$,

$$Z_v := \{x \in \mathbb{R}^d : G(x, v) \in A\}. \tag{16}$$

The randomized depth area at level τ is defined as the random set Z_V.

b) Similarly, define the empirical randomized depth region as $Z_{V,n}$, where

$$Z_{v,n} := \{x \in \mathbb{R}^d : G_n(x, v) \in A_n\}. \tag{17}$$

These randomized depth areas are random subsets of $\mathbb{R}^d$ of exact mass τ:

Proposition 4. *The randomized and empirical randomized depth areas Z_V and $Z_{V,n}$ are of exact mass τ, i.e.,*

$$\mathrm{P}^*(X \in Z_V) = \tau \text{ and } \mathrm{P}^*(X_n^* \in Z_{V,n}) = \tau.$$

Proof. By definition, we get

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

Similarly, $\mathrm{P}^*(X_n^* \in Z_{V,n}) = \mathrm{P}^*(G_n(X_n^*, V) \in A_n) = \mathrm{P}^*(U_n \in A_n) = \tau$. □

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

Theorem 5. *If i) P^X is discrete, or ii) $\mathrm{P}^X \ll \lambda^d$ and (13) holds, then the empirical randomized depth area $Z_{V,n}$ is asymptotically consistent in the P^X-symmetric difference distance towards Z_V, viz.*

$$\mathrm{P}^X(Z_{V,n} \triangle Z_V) \to 0, \text{ as } n \to \infty,$$

with P*-probability one.*

The membership depth functions z, z_n have a close connection with the randomized sets $Z_V, Z_{V,n}$:

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

and similarly for z_n,

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

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

Corollary 6.

1. *For P^X either absolutely continuous or discrete, $\mathrm{P}^X(z) = \mathrm{P}^{X_n^*}(z_n) = \tau$.*
2. *If i) P^X is discrete, or ii) $\mathrm{P}^X \ll \lambda^d$ and (13) holds, then, $\mathrm{P}^X|z_n - z| \to 0$, with* P*-probability one, as $n \to \infty$.*

Proof.

1. By the law of total expectation,

$$\mathrm{P}^X(z) = \mathrm{E}^*z(X) = \mathrm{P}^*(X \in Z_V) = \tau,$$

and similarly for z_n.

2. By the elementary properties of conditional expectation,

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

by Theorem 5.

□

4. Empirical depth areas obtained from a smoothed empirical measure

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 to obtain empirical random depth sets $Z_{V,n}$. If P^X is continuous, $Z_{V,n}$ is a cloud of the sample points, whereas the population depth set 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 mappings, i.e., degenerate Markov morphisms, and sets are now transformed into sets.

A probabilistic description of the (kernel) smoothing procedure is as follows: On $(\Omega^*, \mathcal{A}^*, \mathrm{P}^*)$ where the r.v.s. live, add to the bootstrap representer X_n^* of the empirical measure a small "scaled error" with multivariate bandwidth h_n from some fixed independent r.v. η with continuous distribution function K, i.e., define

$$\hat{X}_n^* := X_n^* + h_n \eta. \tag{18}$$

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

$$\mathrm{P}^{\hat{X}_n^*} = \mathrm{P}^{X_n^*} * \mathrm{P}^{h_n \eta}.$$

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

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where K is the joint cdf of η. Since $\hat{F}_n$ is continuous, its copula is unique, and one can define its corresponding empirical copula representer $\hat{U}_n$ via the multivariate marginal probability integral transform,

$$\hat{U}_n := \hat{G}_n(\hat{X}_n^*), \tag{19}$$

and denote by $\hat{C}_n$ its (copula) c.d.f. The rest of the procedure is as before: Monge-mass transport $\mathrm{P}^{\hat{U}_n}$ to P^S by the transport map $R_{\hat{C}_n}$ with inverse $Q_{\hat{C}_n}$. Eventually, $\hat{S}_n$ is obtained by setting $\hat{S}_n = R_{\hat{C}_n}(\hat{U}_n)$. One has transformed all corresponding measures by pushing forward them by mappings and so we can reason at the level of random variables according to the diagram in Fig. 4.

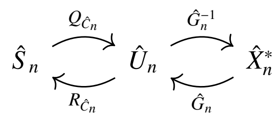

Fig. 4: Transformations of the smooth empirical measure by mappings of the corresponding random variables

For continuous G, the distributional transform $X \to G(X, V)$ reduces to the probability integral transform $X \to G(X)$, so the population counterparts of (18) and (19) are obtained similarly by transformations of random variables by non-randomized mappings, as in the diagram in Fig. 5. (Recall that S, U, X are defined on $(\Omega^*, \mathcal{A}^*, \mathrm{P}^*)$).

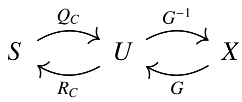

Fig. 5: Transformations of random variables by mappings–population version for a continuous X

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

$$Z = G^{-1}(Q_C(B)), \quad \hat{Z}_n = \hat{G}_n^{-1}(Q_{\hat{C}_n}(B)).$$

Remark 1 (On bandwidth and kernel choice)**.** In (18) it is recommended to choose η with independent components, i.e., a product kernel K, so that one does not introduce artificial dependence in the components of $\hat{X}_n^*$. Moreover, $\hat{U}_n$ in (19) is obtained from $\hat{X}_n^*$ by the transform $\hat{G}_n$ which acts marginal by marginal. Therefore, for the choice of the multivariate bandwidth h_n, one can use univariate bandwidth choice techniques for each component of $\hat{G}_n$ and optimize each marginal bandwidth of h_n separately. The optimal bandwidth choice depends on the degree of regularity of the involved marginal densities and can be effected using nonparametric methods like plug-in or least squares cross-validation, see, e.g., [18] or [17].

The following Proposition is the analogue of the main Theorem 6.2 in Faugeras and Rüschendorf [8]:

Proposition 7. *If* P^X *is absolutely continuous, and* $h_n \downarrow 0$*, one has, with* P *probability one,*

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

In turn, Proposition 7 translates at the level of depth sets as follows:

Corollary 8. *If* P^X *is absolutely continuous, and* $h_n \downarrow 0$*, the* P^X *symmetric distance between the population depth area* Z *and its empirical smoothed counterpart* $\hat{Z}_n$ *becomes asymptotically negligible,*

$$\mathrm{P}^X(\hat{Z}_n \triangle Z) \to 0, \ \mathrm{P}\text{-}a.s.$$

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

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

5. Simulations

We present below some numerical simulations which illustrate the properties of the quantile areas obtained. The proposed depth areas are easily computable thanks to a fast implementation of the optimal transport step using the entropy regularised / Sinkhorn algorithm. Details on the implementation and extensive simulations will be given in a subsequent paper of the authors.

5.1. 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 P^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.

Fig. 6 illustrates Corollary 8 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 Fig. 1 in Hallin et al. [11], one also obtains nicely converging nested regions, as expected, somehow less “spiked”. 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.

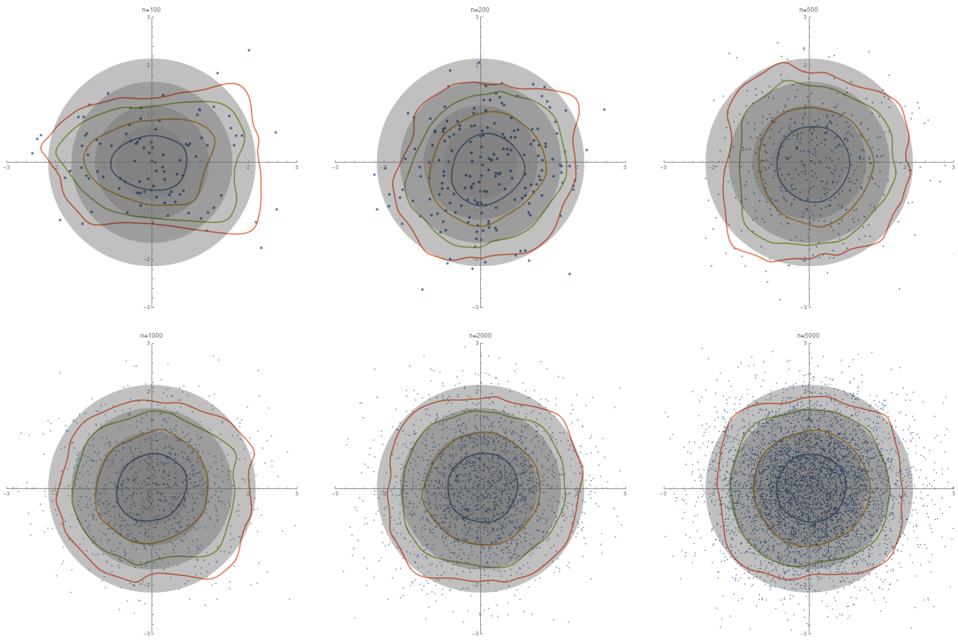

Fig. 6: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 (green), .90 (red)) computed from n = 100, 200, 500, 1000, 2000, 5000 i.i.d. observations from a bivariate standard Gaussian distribution, along with their (spherical) theoretical counterparts (shaded disks).

For the randomized depth areas of Section 3, one obtains as quantile contours a subset of the data points. These can be connected (by linear or cyclical interpolation) to give a contour line. In turn, one gets a similar picture, which is therefore omitted.

5.2. Distributions with non-convex/non-connected support

Fig. 7 and 8 illustrate the remarkable ability of the randomized depth areas of Section 3 to pick the correct geometry of a distribution with non-convex or non-connected support. 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” or disconnected geometry of the distribution considered.

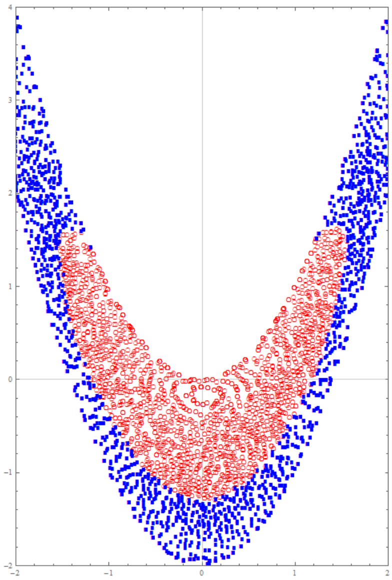

Fig. 7: Non-convex quantile region obtained for the regression model $X = (X_1, X_2)$, $X_2 = -1 + X_1^2 + \epsilon$, $X_1 \sim U_{[-2,2]}$ independent of $\epsilon \sim U_{[-1,1]}$, . (Red) circles: sample points inside the 0.5−quantile region, (Blue) filled squares: sample points sample points outside the 0.5−quantile region. 5000 observations.

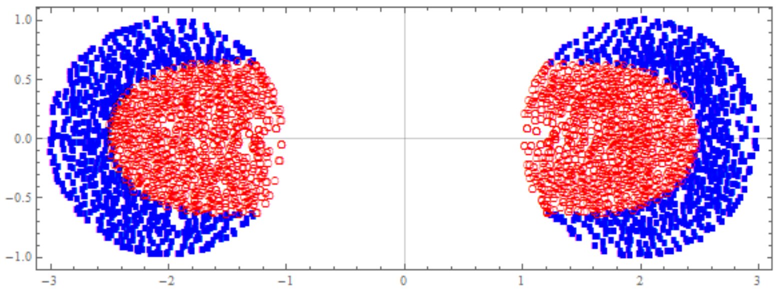

Fig. 8: Non-connected quantile region for a uniform distribution on two disjoint disks. (Red) circles: sample points inside the 0.5−quantile region, (Blue) filled squares: sample points sample points outside the 0.5−quantile region.

5.3. Monotone equivariance

Fig. 9 and 10 illustrate the monotone equivariance property of the proposed quantile areas on a Frank copula model with varied marginals (smoothed approach).

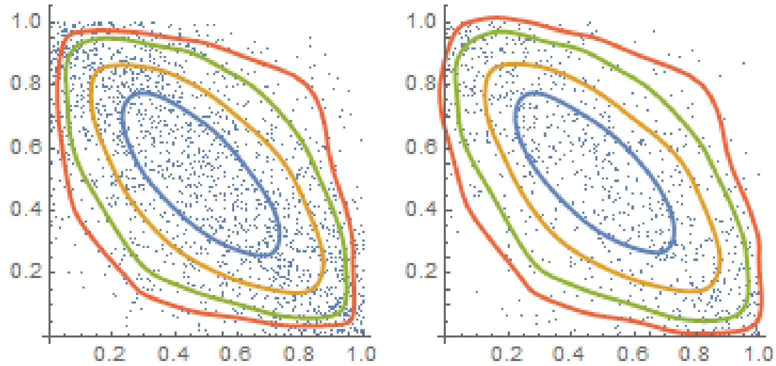

Fig. 9: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 green), .90 red)) computed from $n = 1000$ i.i.d. observations from a Frank ($\theta = 500$) copula distribution with Uniform marginals. left panel: bootstrap sample from the empirical copula and empirical contours in the copula space. right panel: observed data and empirical contours in the original sample space.

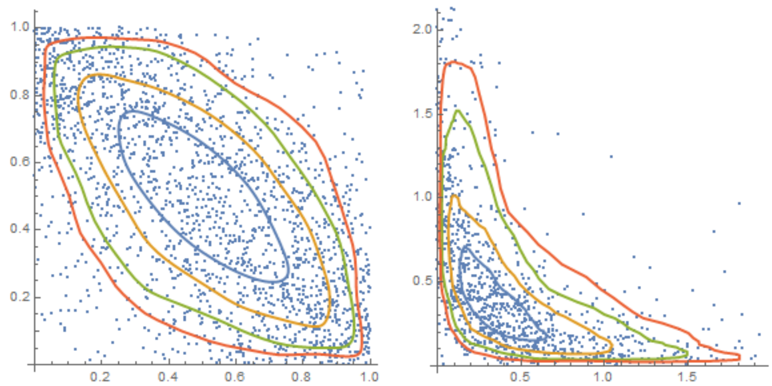

Fig. 10: Smoothed empirical quantile contours (probability contents .25(blue) .50 (brown), .75 green), .90 red)) computed from $n = 1000$ i.i.d. observations from a Frank ($\theta = 500$) copula distribution with Exponentials $Exp(2)$ marginals in the copula space (left) and the original sample space (right).

In Fig. 9, the distribution has uniform marginals and, as expected, one obtains the same quantile contours in the copula space $[0,1]^2$ (left panel) as in the sample space $\mathbb{R}^2$ (right panel). In Fig. 10, both marginals are changed to two exponential $Exp(2)$ distributions: one obtains, up to sample fluctuations, the same contour regions as before in the copula space, which reveals the underlying dependence structure of the model. However, in the observation 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.

Conclusion

We proposed three multivariate notions of central quantile regions: membership functions z, randomized depth areas Z_V, and marginally smoothed depth areas $\hat{Z}$. These are based on the copula and mass transportation approach of Faugeras and Rüschendorf [8] to multivariate quantiles. In particular, their empirical counterparts z_n, $Z_{V,n}$, $\hat{Z}_n$ attain the exact level and are strongly consistent. The introduced depth areas give a valid representation of the central quantile area of a multivariate distribution, and thus are a valuable tool for their analysis. In the following comments, we summarize the main results and point out some interesting aspects of our approach:

- The copula step of the proposed method turns out to have several theoretical advantages: it regularizes the optimal transportation problem between absolutely continuous measures on bounded domains, and so one does not need the regularity assumptions as in Chernozhukov et al. [3], such as existence of moments, compactness and convexity of the support of P^X. This feature is obtained by our particular choice of copula representations, obtained by distributional transforms.

- It is noteworthy to remark that the marginal quantile transformation step, by the distributional transform $G(., V)$ and the marginal quantile function G^{-1} in the opposite direction, can be seen as optimal W_2 transports between P^X and the copula P^U, as follows directly from the basic characterization result for optimal transports in Rüschendorf and Rachev [16].

  The question thus arises to what extent the measures P^S, P^U, P^X are compatible in the sense of composition of optimal transportation maps, as in Definition 2.3.1 in Panaretos and Zemel [14], i.e., to determine whether the proposed copula-marginal two steps approach is equivalent to the direct approach of Chernozhukov et al. [3]. In general, the composition of two successive optimal transportation maps is not the optimal transportation map from the first to the last measure. As a consequence, both approaches lead to different quantile areas.

  In the continuously differentiable case, Panaretos and Zemel [14] p. 50 show that a necessary condition is the commutativity of gradient of the transport maps. In our setting, this corresponds to the commutativity of the matrices $\nabla G^{-1}(Q_C(s))$ and $\nabla Q_C(s)$. As G^{-1} is made of separable variables, ∇G^{-1} is diagonal, which further reduces, in case of identical marginals, to a multiple of the identity matrix. For "approximately compatible" optimal transport maps, one gets constructions looking similar. This explains why Figures 6, resp. 7, give similar quantile areas as the direct approach, Figure 1 in Hallin et al. [11], resp. Figure 2 in Chernozhukov et al. [3].

- A relevant practical advantage of the proposed two steps approach is that for changing marginal conditions, the quantile based areas remain the same at the copula level, while the quantile area at the observational level can easily be adapted by the marginal quantile step. As a consequence, one can obtain an overview in the change of quantile regions w.r.t. a change on marginals. In comparison, the direct approach of Chernozhukov et al. [3] and Hallin et al. [11] requires for each marginal situation a separate optimal transportation problem to be solved. This point seems quite relevant for prediction purposes, e.g. in risk analysis in order to predict changes in the quantile and tail regions under different scenarii of marginals.

- Because copulas are invariant w.r.t. monotone increasing transformations of the marginals, see, e.g., Theorem 2.4.3 in Nelsen [13], one obtains depth areas which are equivariant w.r.t. monotone increasing transformations of the marginals of X. In other words, if X is transformed into $T(X) := (T_1(X_1), \dots, T_d(X_d))$, where each map $T_1, \dots, T_d$ are increasing, the depth areas will remain the same at the copula level, and will be transformed at the X level by the corresponding T. One therefore retains a key feature of univariate quantiles, equivariance w.r.t. a nonlinear monotone change of scale. This monotone marginal equivariance is a useful property as regards robustness of the quantile areas w.r.t. general (nonlinear) changes of scales. It is different from and complements the affine equivariance property of the Chernozhukov et al. [3] direct approach. When the above-mentioned compatibility condition holds, the two steps approach will also inherit the affine equivariance property of the direct approach.

- As may appear surprising at first sight, the Markov morphism view on multivariate quantiles we advocate in Faugeras and Rüschendorf [8] and in this paper allows to define multivariate quantile objects even for a discrete distribution P^X! This increased generality requires a conceptual change in the corresponding notion of "multivariate central quantile areas": one either can introduce membership *functions* z, (which can be interpreted as describing the probability of membership), or consider *randomized* quantile sets Z_V. This is also convenient for their statistical counterparts z_n and $Z_{V,n}$, which leads to the problem of making inference from the discrete empirical measure $P^{X_n^*}$.

- The distributional transform approach to copulas allows to obtain genuine absolutely continuous empirical copulas, see Faugeras [6, 7]. This results in a nonparametric procedure without any bandwidth to optimize. In particular, the empirical membership function z_n and randomized depth area $Z_{V,n}$ are bandwidth free. For the marginally smoothed quantile areas $\hat{Z}_n$, 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. [3], 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.

Acknowledgments

We thank the Editor and referees for their constructive comments. Olivier P. Faugeras acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

6. Appendix: Proofs

6.1. Absolute continuity and open support set of the chosen copula measure.

In this section, we state and an auxiliary result on the chosen copula measure P^U, obtained by the distributional transform (5), with uniform randomizer $V \sim \lambda^d$. Under the assumption that P^X is discrete or that P^X is absolutely continuous together with the mild regularity assumption (13) on the support, Lemma 9 below shows the absolute continuity of P^U, together with the existence of an open support set of P^U.

If G is continuous, the copula is unique, and P^U is the push-forward of P^X by G. If, in addition, P^X is absolutely continuous, it is not immediately clear that P^U is also absolutely continuous, since G may have flat spots (the result would be obvious if the marginals of X had strictly positive densities). Intuitively, the points where the components of G have a zero derivative are those which are not charged by P^X.

If P^X is discrete, its associated copulas are no longer unique. They should not be considered as a given, but as a construct, see Faugeras [7] for details. In particular, one can always obtain an absolutely continuous copula by the specific distributional transform construct with an absolutely continuous randomizer. This is the standard construction in Sklar's Theorem, which is also absolutely continuous in the discrete case, as is shown below.

Lemma 9. *Let $U = G(X, V)$ be the distributional transform (5), with independent randomizer $V \sim \lambda^d$. If i) P^X is discrete, or ii) $\mathrm{P}^X \ll \lambda^d$ and (13) holds, then $\mathrm{P}^U \ll \lambda^d$ and there exists an open support set of P^U, i.e., there exists an open subset W s.t. $\mathrm{P}^U(W) = 1$. In particular, for case ii), the copula density is given for $u \in W$ by*

$$c(u) = \frac{f(G^{-1}(u))}{\prod_{i=1}^d g_i(G_i^{-1}(u_i))}, \tag{20}$$

where f denotes the density of P^X, while g_i denotes the marginal density of the ith component P^{X_i}.

Proof. i) Case one: P^X discrete. By conditioning on $X = x$, the distribution P^U writes as a mixture $\mathrm{P}^U = \sum_x \mathrm{P}(X = x)\mathrm{P}^{G(x,V)}$, since V is chosen independent of X. In addition, since $V \sim \lambda^d$, $\mathrm{P}^{U|X=x} = \mathrm{P}^{G(x,V)}$ is uniform on the cube $[G(x-), G(x)]$, a fortiori absolutely continuous. Hence, $\mathrm{P}^U \ll \lambda^d$. Moreover, it is clear that

$$W := \bigcup_{x:\mathrm{P}^X(x)>0}]G(x-), G(x)[$$

is an open support set of P^U.

ii) Case two: $\mathrm{P}^X \ll \lambda^d$ and (13) holds. The proof is based on the transformation formula Theorem 7.26 in Rudin [15] p. 154: for every measurable function $h : \mathbb{R}^d \mapsto [0, \infty]$, one has

$$\int_{T(N)} h d\lambda^d = \int_N (h \circ T)|J_T| d\lambda^d, \tag{21}$$

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

Since G is continuous, the distributional transform reduces to the probability integral transform $U = G(X)$. Let A be any Borel set of $[0, 1]^d$. Let $E = \{x \in \mathbb{R}^d : f(x) > 0\}$ and $M = \{x \in \mathbb{R}^d : \prod_{i=1}^d g_i(x_i) > 0\}$. If $x \in M^c$, then

$\exists i \in \{1,\dots,d\}$ s.t. $0 = g_i(x_i) = \int f(x)d\lambda_{-i}^{d-1}$, where the integration is w.r.t. all coordinates $x_1,\dots,x_d$, except x_i. Therefore, $f(x) = 0$ a.e. Hence, $E \subset M$, λ^d−a.e. and $\mathring{E} \subset \mathring{M}$, λ^d−a.e.

Moreover, on $\mathring{M}$, G is continuous and strictly increasing (coordinate wise) with continuous inverse G^{-1}. Therefore, $G(\mathring{E} \cap \mathring{M})$ is open.

Thus, one has

$$\begin{aligned}
\mathrm{P}^U(A) &= \mathrm{P}^{G(X)}(A) = \int \mathbb{1}_A(G(x))\mathrm{P}^X(dx) = \int \mathbb{1}_A(G(x))\mathbb{1}_E(x)f(x)\lambda^d(dx) \\
&= \int \mathbb{1}_A(G(x))\mathbb{1}_{\mathring{E}}(x)f(x)\lambda^d(dx) = \int \mathbb{1}_A(G(x))\mathbb{1}_{\mathring{E}\cap\mathring{M}}(x)f(x)\lambda^d(dx) \qquad (22)\\
&= \int_{G^{-1}(G(\mathring{E}\cap\mathring{M}))} \mathbb{1}_A(G(x))f(x)\lambda^d(dx), \qquad (23)
\end{aligned}$$

where (22) and (23) follow from assumption (13), $\mathring{E} \subset \mathring{M}$, λ^d−a.e. and $G^{-1} \circ G = id$ on $\mathring{M}$. On $G(\mathring{M})$, G^{-1} is differentiable, one-to-one, with inverse G continuous, and the Jacobian of G^{-1} writes as $|J_{G^{-1}}(u)| = \left(\prod_{i=1}^d g_i(G_i^{-1}(u_i))\right)^{-1}$. We can now apply Rudin's transformation formula (21), with $O = N = G(\mathring{E} \cap \mathring{M})$, $T = G^{-1}$. Thus, (23) becomes

$$\mathrm{P}^U(A) = \int_{A\cap G(\mathring{E}\cap\mathring{M})} \frac{f(G^{-1}(u))}{\prod_{i=1}^d g_i(G_i^{-1}(u_i))}\lambda^d(du). \qquad (24)$$

Equation (24) means that $\mathrm{P}^U \ll \lambda^d$ has the copula density $c(u) = f(G^{-1}(u))/(\prod_{i=1}^d g_i(G_i^{-1}(u_i)))\mathbb{1}_{G(\mathring{E}\cap\mathring{M})}(u)$. On $G(\mathring{E} \cap \mathring{M})$, $0 < c(u) < \infty$. Moreover, for $A = G(\mathring{E} \cap \mathring{M})$ in (24) and (22), one has, under the stated assumptions, that

$$\mathrm{P}^U(G(\mathring{E} \cap \mathring{M})) = \mathrm{P}^X(G^{-1} \circ G(\mathring{E} \cap \mathring{M})) = P^X(\mathring{E} \cap \mathring{M}) = \mathrm{P}^X(\mathring{E}) = 1.$$

Therefore $W := G(\mathring{E} \cap \mathring{M})$ is an open support set of P^{U}.

□

Remark 2.

i) Assumption (13) is satisfied if E is open (i.e., $f = 0$ on the boundary $\mathrm{bd}(E)$ of the support), or if $\lambda^d(\mathrm{bd}(E)) = 0$. The latter condition is satisfied, e.g., if $\mathrm{bd}(E)$ has Hausdorff dimension $< d$. A counter-example in dimension two would be a distribution whose boundary of the support is an Osgood curve. Absolutely continuous distributions whose boundary of the support has positive λ^d Lebesgue measure are somewhat pathological and are seldom met in practice. Hence, the regularity assumption (13) is very mild and appears natural.

ii) For a mixture of absolutely continuous and discrete distributions, i.e., for $\mathrm{P}^X = \alpha\mathrm{P}^{X^{ac}} + (1-\alpha)\mathrm{P}^{X^{dis}}$, with $0 < \alpha < 1$, $\mathrm{P}^{X^{ac}} \ll \lambda^d$, $\mathrm{P}^{X^{dis}}$ discrete, one can construct a corresponding absolutely continuous copula measure, with a modified distributional transform $\tilde{U}$ as follows: X writes $X = \zeta X^{ac} + (1-\zeta)X^{dis}$, where $\zeta \sim B(1,\alpha)$ and ζ, X^{ac}, X^{dis} are independent. Define $\tilde{U} = \zeta G^{ac}(X^{ac}) + (1-\zeta)G^{dis}(X^{dis}, V)$, where G^{ac} is the vector of marginal c.d.f. of X^{ac} and $G^{dis}(X^{dis}, V)$ is the distributional transform of X^{dis}. Then, the c.d.f. of $\tilde{U}$ is a copula associated to X. Its distribution is $\mathrm{P}^{\tilde{U}} = \alpha\mathrm{P}^{G^{ac}(X^{ac})} + (1-\alpha)\mathrm{P}^{G^{dis}(X^{dis},V)}$. Hence, $\mathrm{P}^{\tilde{U}}$ is a mixture of two absolutely continuous distributions by cases one and two of Lemma 9, and thus is absolutely continuous.
However, the empirical copula representer $U_n = G_n(X_n^*, V)$ will not converge to this modified $\tilde{U}$. One would require a different statistical setting (separate samples for each component) to construct a converging empirical copula representer. Therefore, we do not include this case in our results.

6.2. Proofs of the main results

Proof of Theorem 2.

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

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

and similarly

$$\mathrm{P}^{U_n}(A_n) = \mathrm{P}^{U_n}(a_n) = \mathrm{P}^{U_n}(\mathcal{R}_{C_n} b) = (\mathrm{P}^{U_n}\mathcal{R}_{C_n})(b) = \mathrm{P}^S(b) = \tau.$$

(b) $|\mathrm{P}^U(a_n - a)| \leq \mathrm{P}^U|a_n - a| \to 0$, as $n \to \infty$ by Theorem 2 1. (c).

(c) One has

$$\mathrm{P}^U(|a_n - a|) = \mathrm{P}^U(|\mathbb{1}_{A_n} - \mathbb{1}_A|) = \mathrm{P}^*(U \in A, U \notin A_n) + \mathrm{P}^*(U \notin A, U \in A_n) := (I) + (II)$$

with

$$(I) \quad = \quad \mathrm{P}^*(U \in A, U \notin A_n, U_n \in A_n) + \mathrm{P}^*(U \in A, U \notin A_n, U_n \notin A_n),$$

where U, U_n are the coupling constructions of Theorem 6.2 in Faugeras and Rüschendorf [8]. Let d be the Euclidean distance on $\mathbb{R}^d$. By the triangle inequality,

$$d(U, A_n) \leq d(U, U_n) + d(U_n, A_n).$$

By Theorem 6.2 in Faugeras and Rüschendorf [8], $d(U, U_n) \overset{\mathrm{P}^* a.s.}{\longrightarrow} 0$, and if $U_n \in A_n$ then $d(U_n, A_n) = 0$. Therefore, if $U_n \in A_n$, then $d(U, A_n) \to 0$, i.e., $U \in A_n$ asymptotically with probability one. Hence, $\mathrm{P}^*(U \in A, U \notin A_n, U_n \in A_n) \to 0$. Similarly, $\mathrm{P}^*(U \in A, U \notin A_n, U_n \notin A_n) \leq \mathrm{P}^*(S \in B, S_n \notin B)$. By the triangle inequality,

$$d(S_n, B) \leq d(S_n, S) + d(S, B).$$

By Theorem 6.2 in Faugeras and Rüschendorf [8], $d(S, S_n) \to 0$ and $S \in B$ implies $S_n \in B$ asymptotically with probability one. Hence, $\mathrm{P}^*(S \in B, S_n \notin B) \to 0$. 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.

$$\begin{aligned} S \in B \quad &\Leftrightarrow \quad R_C(U) \in B \Leftrightarrow U \in R_C^{-1}(B) = Q_C(B) = A \\ &\Rightarrow \quad G^{-1}(U) \in G^{-1}(A) = Z \Leftrightarrow X \in Z, \end{aligned} \tag{25}$$

where $X \sim F$ also sits on the probability space $(\Omega^*, \mathcal{A}^*, \mathrm{P}^*)$ of Theorem 6.2 in Faugeras and Rüschendorf [8], and (25) follows from the fact that $G^{-1}(U) = X$, P^*-a.s., by definition of the distributional transform. Therefore, $\tau = \mathrm{P}(S \in B) \leq \mathrm{P}(X \in Z)$. The proof for Z_n is similar: for $S_n = R_{C_n}(U_n)$,

$$S_n \in B \Leftrightarrow U_n \in A_n \Rightarrow X_n^* \in Z_n. \tag{26}$$

The proof that $\mathrm{P}^X(z) = \mathrm{P}^{X_n^*}(z_n) = \tau$ follows as in Theorem 2 1. (a) from the definitions of the Markov morphisms. If P^X is continuous, the distributional transform $X \to G(X, V)$ reduces to $X \to G(X)$ and the implications in (25) become equivalences P^* a.s.,

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

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

(b) One has

$$\mathrm{P}^X(|\mathbb{1}_{Z_n} - \mathbb{1}_Z|) = \mathrm{P}^*(X \in Z, X \notin Z_n) + \mathrm{P}^*(X \notin Z, X \in Z_n) := (I) + (II)$$

with

$$(I) = \mathrm{P}^*(X \in Z, X \notin Z_n, U_n \in A_n) + \mathrm{P}^*(X \in Z, X \notin Z_n, U_n \notin A_n) := (Ia) + (Ib).$$

By the implication in (26), one has

$$(Ia) := \mathrm{P}^*(X \in Z, X \notin Z_n, U_n \in A_n) \leq \mathrm{P}^*(X \in Z, X \notin Z_n, X_n^* \in Z_n) \leq \mathrm{P}^*(X \notin Z_n, X_n^* \in Z_n)$$

The triangle inequality gives $d(X, Z_n) \leq d(X, X_n^*) + d(X_n^*, Z_n)$. By Theorem 6.2 in Faugeras and Rüschendorf [8], $d(X, X_n^*) \to 0$ as $n \to \infty$, and $X_n^* \in Z_n$ implies $d(X^*, Z_n) \to 0$. Therefore, $(Ia) \to 0$. For P^X continuous, by the equivalences in (26) and (27),

$$(Ib) := \mathrm{P}^*(X \in Z, X \notin Z_n, U_n \notin A_n) \leq \mathrm{P}^*(S \in B, S_n \notin B).$$

The triangle inequality gives $d(S_n, B) \leq d(S_n, S) + d(S, B)$. By Theorem 6.2 in Faugeras and Rüschendorf [8], $d(S_n, S) \to 0$ as $n \to \infty$, and $S \in B$ implies $d(S_n, B) \to 0$. Therefore, $(Ib) \to 0$. The treatment of (II) is similar. $\mathrm{P}^X(Z_n) \to \tau$ follows.

(c) See Corollary 6.

□

Proof of Theorem 5. One has, with P and P*-probability one,

$$\begin{aligned}\{X \in Z_V\} &= \{G(X,V) \in A\} = \{U \in A\} = \{S \in B\}\\ \{X \in Z_{V,n}\} &= \{G_n(X,V) \in A_n\} = \{G_n(X,V) \in Q_{C_n}(B)\} = \{R_{C_n} \circ G_n(X,V) \in B\}.\end{aligned}$$

Therefore,

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

It suffices to show that

$$R_{C_n}(G_n(X,V)) \xrightarrow{\mathrm{P}^*} S, \tag{28}$$

where $\xrightarrow{\mathrm{P}^*}$ stands for convergence in P*-probability. Indeed, if (28) holds, then this imply convergence in distribution and thus

$$\mathrm{P}^*(X \in Z_V \setminus Z_{V,n}) = \mathrm{P}^*(\{S \in B\} \setminus \{R_{C_n} \circ G_n(X,V) \in B\}) \to 0,$$

since B is a continuity set of P^S. One has similarly, for the other inclusion,

$$\mathrm{P}^*(X \in Z_{V,n} \setminus Z_V) \to 0,$$

which yields the result.

Let us show (28). By ergodicity, for all $x \in \mathbb{R}^d$, with P-probability one,

$$|G_n(x_-) - G(x_-)| \to 0, \quad |G_n(x) - G(x)| \to 0.$$

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

$$G_n(X,V) \to G(X,V) = U, \tag{29}$$

with P and P*-probability one.

Consider next the decomposition

$$\begin{aligned}R_{C_n} \circ G_n(X,V) - S &= R_{C_n} \circ G_n(X,V) - R_C \circ G_n(X,V) + R_C \circ G_n(X,V) - R_C \circ G(X,V)\\ &:= (I) + (II)\end{aligned} \tag{30}$$

Convergence of (I) in (30): For i) P^X discrete, or ii) P^X absolutely continuous and assumption (13), Lemma 9 in Section 6.1 entails that there exists bounded open sets $W, W_n \subset [0,1]^d$, s.t. $\mathrm{P}^U(W) = 1$ and $\mathrm{P}^{U_n}(W_n) = 1$. Let ϕ be the optimal convex potential associated with the optimal transport of P^U towards P^S. Let Σ be the λ^d-null set which the subdifferential $\partial\phi$ is not univalued. Let $\epsilon > 0$. By inner regularity of the Lebesgue measure, and since $\mathrm{P}^{\mathrm{U}} \ll \lambda^d$, there exists a compact set $K \subset W \setminus \Sigma$, s.t.

$$\mathrm{P}^{\mathrm{U}}(K^{\mathrm{c}}) < \epsilon. \tag{31}$$

Decompose (I) further as,

$$(I) = (R_{C_n} \circ G_n(X,V) - R_C \circ G_n(X,V))\,\mathbb{1}_{G_n(X,V)\in K} + (R_{C_n} \circ G_n(X,V) - R_C \circ G_n(X,V))\,\mathbb{1}_{G_n(X,V)\notin K} := (Ia) + (Ib).$$

By Theorem 6.2 in Faugeras and Rüschendorf [8], $d(U, U_n) \xrightarrow{\mathrm{P}^* a.s.} 0$, therefore by the Portmanteau Theorem,

$$\liminf \mathrm{P}^*(U_n \in W) \geq \mathrm{P}^*(U \in W) = 1.$$

Thus, $P^*(U_n \in W \cap W_n) \to 1$. Hence, the restriction $\mu_n := P^{U_n}|_{W\cap W_n}$ converges in distribution to $\mu := P^U$, and similarly, $\nu_n := R_{C_n}\#\mu_n \xrightarrow{d} \nu := P^S$.

To deal with (Ia), we apply Proposition 1.7.11 of Panaretos and Zemel [14]: since P^{U_n}, P^S live on bounded sets and $P^U \ll \lambda^d$, their assumption 1 is satisfied for $\mu_n \xrightarrow{d} \mu$ and $\nu_n \xrightarrow{d} \nu$. Therefore, R_{C_n} converges uniformly on any compact set of Lebesgue density points of the support of P^U on which the subdifferential $\partial\phi$ is univalued. But $\partial\phi = \{R_C\}$ is univalent on $W \setminus \Sigma$, since $P^U \ll \lambda^d$. Therefore, $\sup_{y\in K} \|R_{C_n}(y) - R_C(y)\| \to 0$, which entails that $(Ia) \to 0$ a.s.

To deal with (Ib), by the Portmanteau theorem and (31),

$$\limsup_{n\to\infty} P^*(G_n(X, V) \notin K) \leq P^*(U \notin K) \leq \epsilon.$$

Moreover, since $R_C, R_{C_n} \in B(0, 1)$, $\|(Ib)\| \leq 2\mathbb{1}_{G_n(X,V)\notin K}$. For every $\epsilon_1 > 0$, by Markov inequality,

$$P^*(\|(Ib)\| > \epsilon_1) \leq 2\frac{P^*(G_n(X, V) \notin K)}{\epsilon_1} \leq 2\epsilon/\epsilon_1$$

for n large enough. Since ϵ is arbitrary, letting $\epsilon \downarrow 0$ yields $(Ib) \xrightarrow{P^*} 0$.
Convergence of (II) in (30): By Proposition 2.4 in Cuesta-Albertos et al. [4], R_C is λ^d-a.e. continuous. Therefore, by the continuous mapping Theorem, (29) yields

$$(II) = R_C \circ G_n(X, V) - R_C \circ G(X, V) \xrightarrow{P^* a.s.} 0.$$

Collecting all elements, we have thus shown that $R_{C_n}(G_n(X, V)) \xrightarrow{P^*} S$, which ends the proof.

□

Remark 3.

i) The proof is based on the aforementioned general result of uniform convergence on compacta of optimal transportation maps of Panaretos and Zemel [14], Proposition 1.7.11. The latter does not require a convexity assumption on the support of the measures, nor a homeomophism condition of the transport maps, as mandated by previous results in the literature, such as Theorems A.1 and A.2 in Chernozhukov et al. [3] (see also Theorem 1.7.7 in Panaretos and Zemel [14]). The second main ingredient of the proof is the continuity result for optimal mappings of Cuesta-Albertos et al. [4].

ii) As in Chernozhukov et al. [3] and Hallin et al. [11], in our approach we take as source measure P^S the spherical measure defined as the product of the uniform distributions on $[0, 1]$ and on the unit sphere. This choice introduces a complication in the regularity theory of the transport maps, as the density of P^S is unbounded at the center zero. Chernozhukov et al. [3] require a homeomorphism condition (C) for the validity of Theorem A.2, which is studied in Figalli [9] and del Barrio et al. [1].
Alternatively, it would be possible to avoid this difficulty by replacing the spherical P^S by, e.g., the uniform distribution on the unit ball: the latter is also absolutely continuous, with the same bounded convex support, but without a singularity in its density. The only change is that the depth sets of level τ for P^S are no longer the balls B_τ, but the balls of radius $r(\tau)$ s.t.

$$\frac{\pi^{d/2}}{\Gamma(d/2 + 1)}(r(\tau))^d = \tau,$$

i.e., s.t. $P^S(B_{r(\tau)}) = \tau$, as before.

Proof of Proposition 7. The proof follows the main arguments of Theorem 6.2 in Faugeras and Rüschendorf [8], with minor modifications. On the auxiliary probability space $(\Omega^*, \mathcal{A}^*, P^*)$ on which (12) holds, $h_n \downarrow 0$ and (12) in (18) yields, with P-probability one,

$$\hat{X}^*_n \xrightarrow{P^* a.s.} X. \tag{32}$$

We have the decomposition,

$$\hat{U}_n - U = \hat{G}_n(\hat{X}^*_n) - G(\hat{X}^*_n) + G(\hat{X}^*_n) - G(X) \leq \|\hat{G}_n - G\|_\infty + G(\hat{X}^*_n) - G(X), \tag{33}$$

where the operations and the $||.||_\infty$ norm are to be understood componentwise. By ergodicity, one has that $\forall x \in \mathbb{R}^d$, $G_n(x) \to G(x)$. Since G is continuous, this implies by Polya's Theorem, a Glivenko-Cantelli type Theorem for G_n, i.e., $||G_n - G||_\infty \to 0$, with P-probability one. Obviously $\mathrm{P}^{h_n\eta} \stackrel{d}{\to} \delta_0$, so one has, as in Winter [19] Theorem 2, that

$$||\hat{G}_n - G||_\infty \to 0, \tag{34}$$

with P-probability one. Hence, (34) in (33), and continuity of G with (32) entails $\hat{U}_n \stackrel{\mathrm{P}^* a.s.}{\longrightarrow} U$, with P-probability one. The rest of the proof is as Theorem 6.2 in Faugeras and Rüschendorf [8]: by Cuesta-Albertos et al. [4] Theorem 3.4, one gets $\hat{S}_n \stackrel{\mathrm{P}^* a.s.}{\longrightarrow} S$, with P-probability one. □

References

[1] E. del Barrio, A. González-Sanz, M. Hallin, A note on the regularity of optimal-transport-based center-outward distribution and quantile functions, J. Multivariate Anal. 180 (2020) 104671, 13 pp.
[2] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (1991) 375–417.
[3] V. Chernozhukov, A. Galichon, M. Hallin, M. Henry, Monge-Kantorovich depth, quantiles, ranks and signs, Ann. Stat. 45 (2017) 223–256.
[4] J. A. Cuesta-Albertos, C. Matrán, A. Tuero-Díaz, Optimal transportation plans and convergence in distribution, J. Multivariate Anal. 60 (1997) 72–83.
[5] G. Easton, R. McCulloch, A multivariate generalization of quantile–quantile plots, J. Amer. Statist. Assoc. 85 (1990) 376–386.
[6] O. P. Faugeras, Maximal coupling of empirical copulas for discrete vectors, Journal of Multivariate Analysis 137 (2015) 179–186.
[7] O. P. Faugeras, Inference for copula modeling of discrete data: a cautionary tale and some facts, Depend. Model. 5 (2017) 121–132.
[8] O. P. Faugeras, L. Rüschendorf, Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles, Mathematica Applicanda 45 (2017) 3–45.
[9] A. Figalli, On the continuity of center-outward distribution and quantile functions, Nonlinear Anal. 177 (2018) 413–421.
[10] P. Ghosal, B. Sen, Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing, arXiv preprint arXiv:1905.05340 (2019).
[11] M. Hallin, E. del Barrio, J. Cuesta-Albertos, C. Matrán, Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach, Ann. Stat. 49 (2021) 1139 – 1165.
[12] R. J. McCann, Existence and uniqueness of monotone measure-preserving maps, Duke Math. J. 80 (1995) 309–323.
[13] R. B. Nelsen, An Introduction to Copulas, Springer Series in Statistics, Springer, New York, second edition, 2006.
[14] V. M. Panaretos, Y. Zemel, An Invitation to Statistics in Wasserstein Space, SpringerBriefs in Probability and Mathematical Statistics, Springer International Publishing, 2020.
[15] W. Rudin, Real and Complex Analysis, McGraw-Hill Book Co., New York, third edition, 1987.
[16] L. Rüschendorf, S. T. Rachev, A characterization of random variables with minimum L^2-distance, J. Multivariate Anal. 32 (1990) 48–54.
[17] M. P. Wand, M. C. Jones, Kernel Smoothing, volume 60 of Monographs on Statistics and Applied Probability, Chapman and Hall, Ltd., London, 1995.
[18] L. Wasserman, All of Nonparametric Statistics, Springer Texts in Statistics, Springer, New York, 2006.
[19] B. B. Winter, Strong uniform consistency of integrals of density estimators, Canad. J. Statist. 1 (1973) 247–253.