Quantiles as Markov morphisms: a copula and mass transportation approach

Olivier P. Faugeras,
Toulouse School of Economics,
University of Toulouse Capitole, Toulouse, France.
Ludger Rüschendorf,
Abteilung für Mathematische Stochastik,
Albert-Ludwigs University of Freiburg, Freiburg, Germany.

December 13, 2016

Abstract

The objective of this paper is two-fold: the first part (sections 1 to 3) is a discussion and an elucidation at the conceptual level of the notion of multivariate quantile. We argue that the concept of quantile should not be considered as a function but as a Markov kernel from a reference distribution to the considered one. We organize our discussion in three stages of increasing conceptual generality. In section 1, we adopt an analytical point of view: we review the properties of univariate quantile functions (q.f.) and briefly summarize the different approaches considered in the literature to define multivariate q.f. and the related notion of depth. In section 2, we discuss how q.f. and cumulative distribution functions (c.d.f.) arises naturally as reciprocal (randomized) transformations of random variables. We show similarly how copula and conditional q.f./c.d.f. can be viewed from this probabilistic viewpoint. In section 3, we eventually take the final conceptual step and argue, on abstract algebraic grounds, that the object quantile should be regarded at the categorical level of a Markov morphism between probability measures, compatible with some algebraic, ordering and topological structures.

In a second part (sections 4 to 5), we intent to show that the above conceptual discussion can be concretized by proposing a multivariate quantile Markov morphism which combines the copula view and the mass transportation view, elaborating on the recent article by [5]. The proposed Markov morphism is the composition of a copula transformation which, although a random transformation, leaves invariant the dependence structure while regularising the distribution, and a Monge transform arising from a mass transportation problem between a reference spherical measure and the copula measure. The proofs of

the consistency of the empirical version of the proposed quantile to its population version are deferred in section 5.

1 Quantile as a function: a discussion of the analytical view in the literature

Notation

- Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ a measurable Polish space endowed with its Borel sigma algebra $\mathcal{B}(\mathcal{X})$;
- $\mathcal{F}(\mathcal{X})$ stands for $\mathcal{B}(\mathcal{X})$-measurable real-valued functions $f_{\mathcal{X}} : \mathcal{X} \to \mathbb{R}$;
- $\mathcal{P}(\mathcal{X})$ stands for the set of Borel Probability measures P^X on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$;
- Unless specified otherwise, we will work in practice on the Euclidean measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Denote vectors $\mathbf{X}$ by bold letters, and interpret operations between vectors componentwise. $P^{\mathbf{X}}$ will stand for the probability measure associated with its representing variable $\mathbf{X}$.

Let's briefly recall some basic facts about univariate quantile functions (q.f.), which, although elementary, will help to motivate the approaches of sections 2 and 3.

1.1 Univariate quantile functions as inverse functions

Let $X : \Omega \to \mathbb{R}$ be an univariate real r.v. and denote by P^X its corresponding law. The probability measure P^X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ can be characterized analytically by its cumulative distribution function (c.d.f.) $F_X(x) := P^X((-\infty, x])$, (see any probability book and e.g. Szekli [55] for other analytical characterisations). The classical textbook view on the q.f. Q_X of F_X is usually to define it as the left-continuous generalised inverse function of F_X,

$$Q_X(t) = F_X^{\leftarrow}(t) := F_X^{-1}(t) := \inf\{x \in \mathbb{R} : F_X(x) \geq t\}, \quad 0 < t < 1. \qquad (1)$$

An informal rationale for such an "inverse" view could be the following: if X is, as in insurance theory, thought as a random positive monetary quantity which stands for the loss incurred by an insurer, the "risk" carried by X can be approached via two dual paths:

- for a given level x, what is the degree of occurrence that the random loss X be larger than x ? This is quantified by the tail or survival function $\overline{F}_X(x) := P(X > x)$ (or equivalently by the c.d.f. $F_X(x) := P(X \leq x) = 1 - \overline{F}_X(x)$);
- for a given degree of occurrence t, what is the value x_t such that the insurer has probability (at least) t that he will not lose more than x_t ? This is quantified by the q.f. $x_t := Q_X(t)$.

More formally, definition (1) entails that the c.d.f. F_X and q.f. Q_X are in a sort of "inverse duality": for $0 < t < 1$ and $x \in \mathbb{R}$,

$$F_X(x) \geq t \Leftrightarrow x \geq Q_X(t), \tag{2}$$

which entails,

$$F_X(Q_X(t)) \geq t, \text{ and } Q_X(F_X(x)) \leq x. \tag{3}$$

Note, that even in the one-dimensional case, the definition (1) of the q.f. as a left generalised inverse of F is not the sole possibility: one could have chosen as well the right generalised inverse, $F_X^{\rightarrow}(t) := \inf\{x : F(x) > t\}$. Therefore, the choice of a left-continuous inverse for the q.f. and of a right-continuous c.d.f. is a matter of convention, (see e.g. Williams [61] p. 34). The ambiguity in the definition of these generalised inverses comes from the fact that, although the operation $F_X : x \mapsto t := F(x)$ defines a function, the inverse operation $x \leftarrow t$ is an "inverse problem", i.e. $F_X^{-1} : x \leftarrow t$ defines only a *correspondence*, i.e. a multi-valued or set-valued mapping, see Aubin and Frankowska [1] or Rockafellar and Wets [35] for general references on set-valued analysis.

1.2 A summary of some key properties of univariate q.f.

Parzen [25, 26] advocates that it is often advantageous to "think quantile functions" in univariate statistical modeling instead of thinking in terms of c.d.f.:

- q.f. are well-suited for asymptotic inference:
  - they characterize their parent probability measure (so there is no identifiability issues)

    $$P^X = P^Y \Leftrightarrow F_X = F_Y \Leftrightarrow Q_X = Q_Y. \tag{4}$$

  - they are convergence-determining, in the sense that weak and strong convergence can be expressed via q.f. Indeed,
    * univariate q.f. characterizes weak convergence:

      $$F_n \stackrel{d}{\rightarrow} F \Leftrightarrow Q_n \stackrel{d}{\rightarrow} Q, \tag{5}$$

      where $Q_n \stackrel{d}{\rightarrow} Q_n$ stands for convergence in quantile, i.e. $Q_n(t) \rightarrow Q(t)$ at each continuity point t of Q in $(0, 1)$ (See also proposition 7.3.1 p. 112 in [51]);
    * univariate q.f. gives a simple constructive proof of Skorokhod's representation theorem that turns weak convergence into a.s. convergence.
    * Distance between univariate probability measures (Wasserstein's distances) can be expressed via quantile functions.

- univariate q.f. enjoy good invariance properties w.r.t to left-continuous monotone transformations:

  Let $g : \mathbb{R} \mapsto \mathbb{R}$ monotone, left-continuous, $g^{\rightarrow}(y) = \sup\{x \in \mathbb{R} : g(x) \leq y\}$ and $Y = g(X)$.

  - if g monotone non-decreasing, left-continuous, then

$$Q_Y(t) = g(Q_X(t)), \quad \text{and } F_Y(x) = F_X(g^{\rightarrow}(y)); \tag{6}$$

  - if g monotone non-increasing, left-continuous,then $Q_Y(t) = g(Q_X(1-t))$.

- univariate q.f. enjoy good algebraic properties: Gilchrist [14] notices that q.f. can be added and multiplied (when positive);

- Moreover, Parzen [26] argues that univariate q.f. and their empirical version also facilitate the study of order and extreme value distributions: they are the unifying concept behind the notion of confidence intervals, order, ranks, and sign statistics, trimmed means and variances.

1.3 Multivariate quantile functions

If $P^{\mathbf{X}}$ is now a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, it can also be characterized analytically by its multivariate c.d.f. $F(\mathbf{x}) := P^{\mathbf{X}}((-\infty, \mathbf{x}])$, as in the univariate case. Unfortunately, as discussed by e.g. Serfling [46], "the absence of a natural ordering of Euclidean spaces of dimension greater than one [...]" makes the definition of a multivariate q.f. more complicated and diverse. Serfling [46] lists the large literature on the subject and classify several ad-hoc approaches to defining a multivariate q.f. by the type of *method* used to obtain them: methods based on depth functions (method 1), M-estimator based on norm minimization (method 2), Z-estimator of gradients (method 4), inversion of surrogate distributions (method 3), methods based on generalized quantile process (method 5). Refer to [46, 47, 48, 49, 15] for a detailed discussion of the merits and shortcomings of each approach.

These authors favor the "geometric" approach based on depth: In short, the depth $D(\mathbf{x}, F)$ of a point $\mathbf{x} \in \mathbb{R}^d$ with respect to a multivariate cdf F is a measure of the "centrality" of $\mathbf{x}$ w.r.t. to the distribution of mass F, see [47, 48, 46]. The "central point" of maximal depth is a "central point" from which one can define a measure of "outlyingness" and a "center-outward inner region of specified probability", or depth region or *area*,

$$A(\tau, F) := \{\mathbf{x} \in \mathbb{R}^d, D(\mathbf{x}, F) \geq \tau\}, \quad 0 < \tau < 1, \tag{7}$$

whose defining property,

$$P(\mathbf{X} \in A(\tau, F)) \geq \tau,$$

is the multivariate analogue of (3). Depth regions can thus be considered as multivariate extensions of the univariate confidence interval $[Q_X(1-t)/2), Q_X(1-$

$(1-t)/2)]$ of coverage probability t, centered around its median. Serfling and Zuo [47, 49] propose a set of desirable properties depth functions should satisfy and draw some perspectives on quantiles and depths.

However, the classification of q.f. by their methods in Serfling [46], the statement of desirable properties of depth functions in [47, 48], and the perspectives drawn in [49] have a sort of *ad-hoc* character. A *structural* classification of the properties of section 1.2 will be proposed in section 3, once a paradigmatic shift on the subject will have been properly motivated, as we now propose.

2 Quantile as a transformation of random variables: the probabilistic view

In this section, we shift our focus and adopt a probabilistic view on the quantile object. This sort of intermediate point of view between those of sections 1 and 3 will be helpful to motivate the more abstract approach of section 3. It will also allows to view copulas through the probabilistic lens, which will be helpful for the second part of the paper.

2.1 Univariate reciprocal transforms of random variables

Our starting point is that in the univariate case, it is well known that one can transform a r.v. U uniform on $[0,1]$ into a r.v. $X \in \mathbb{R}$ with prescribed c.d.f. F_X, via the *quantile transform* mapping

$$\begin{aligned} Q_X : [0,1] &\rightarrow \mathbb{R} \\ U &\mapsto Q_X(U), \end{aligned}$$

with

$$Q_X(U) \stackrel{d}{=} X, \tag{8}$$

where Q_X is the (left or right) generalised inverse of equation (1). This transformation is key, e.g. to prove (the easy version of) Skorohod's theorem, results on stochastic order, association and a.s. coupling constructions (the method of a single probability space) in classical empirical process theory, (see e.g. Thorisson, chapter one [56], Szekli [55], KMT, Shorack and Wellner, Csorgo and Revesz).

The reciprocal transformation is known as the *Probability integral transformation*

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0,1] \\ X &\mapsto F_X(X). \end{aligned}$$

If F_X is continuous, then

$$F_X(X) \stackrel{d}{=} U \tag{9}$$

However, if F_X is discontinuous, the latter distributional equality is no longer true. Hopefully, define the extended c.d.f.

$$F_X(x, \lambda) := P(X < x) + \lambda P(X = x), \quad \lambda \in [0, 1],$$

and let V a uniform $[0, 1]$ r.v., independent of X. Then, the *distributional transform* is the *randomized* transformation of random variables

$$\begin{array}{rcl} F_X(., V) : \mathbb{R} & \to & [0, 1] \\ X & \mapsto & F_X(X, V) := U \end{array}$$

and is the generalisation of (9) to an arbitrary F: one has, see Rüschendorf [42],

$$U \overset{d}{=} U_{[0,1]}, \quad \text{and} \quad Q_X(U) = X \text{ a.s.} \tag{10}$$

Such a "randomized mapping" $F_X(., V)$ allows to view the pair $(Q_X(.), F_X(., V))$

$$\begin{array}{ccc} U & \overset{Q_X}{\longrightarrow} & X \\ U & \overset{F_X(.,V)}{\longleftarrow} & X \end{array}$$

as genuine reciprocal transformations between r.v.: it bypasses the issue, explained in section 1, of having to represent the inverse operation F_X^{-1} as a multivalued-mapping and even strengthens (8) into an a.s. statement. Of course, the choice of the reference distribution of U, uniform on $[0, 1]$ is conventional. It can be motivated by Laplace's view on randomness: one should generate random variables from an "equiprobable" continuous distribution, viz. a uniform one. It may prove advantageous to use, say, an Exponential or Poisson distribution, as reference distribution and the corresponding transformations then have a different interpretation (in particular, as a hazard function, see Szekli [55]).

2.2 Multivariate transforms of random vectors

For a multivariate $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$, one can similarly look for a transformation

$$\begin{array}{rcl} T : [0, 1] & \to & \mathbb{R}^d \\ U & \mapsto & \mathbf{X} \end{array}$$

from a univariate $U \sim U_{[0,1]}$. Such a generalisation to $X \in \mathbb{R}^d$ or even to $\mathbf{X} \in \mathcal{X}$ a Polish space, would be given by Borel's isomorphism theorem (See Parthasaraty [24], chapter one). Unfortunately, such isomorphisms (which would be perfect candidates for higher-dimensional "quantile functions") are not very convenient tools: no explicit construction, even for $\mathcal{X} = \mathbb{R}^2$, is known; they may be unsmooth and present some pathologies, see [24, 29].

Therefore, it is more convenient to look for a transformation between vectors of same dimensionality, i.e. not from a single univariate U but from a vector $\mathbf{U} = (U_1, \dots, U_d)$,

$$\begin{array}{rcl} [0,1]^d & \leftrightarrows & \mathbb{R}^d \\ \mathbf{U} & \leftrightarrows & \mathbf{X}. \end{array}$$

Moreover, it is expedient for interpretative purposes to impose that the marginals of the reference vector $\mathbf{U}$ have some prescribed distribution, say uniform on $[0,1]$. Basically, there are two competing routes, depending on the dependence structure of the $(U_1, \dots, U_d)$, which leads to either *multivariate quantile representations*, or *copula representations*.

2.2.1 Multivariate Quantile representations

Starting from a vector $\mathbf{U} = (U_1, \dots, U_d)$ of *mutually independent* $U_{[0,1]}$, one wants to generate the vector $\mathbf{X}$ whose distribution is a prescribed c.d.f F.

- The direct transformation $\mathbf{Q}_{\mathbf{U}\to\mathbf{X}} := (Q_1, \dots, Q_d)$,

$$\begin{array}{rcl} \mathbf{Q}_{\mathbf{U}\to\mathbf{X}} : [0,1]^d & \to & \mathbb{R}^d \\ \mathbf{U} & \mapsto & \mathbf{Q}_{\mathbf{U}\to\mathbf{X}}(\mathbf{U}) \end{array} \tag{11}$$

  is the *multivariate conditional quantile transform*, which is the set of successive conditional quantile transforms: set

$$\begin{array}{rcl} Q_1(u_1) & := & Q_{X_1}(u_1) =: x_1, \quad 0 < u_1 < 1, \\ Q_i(u_i|u_{i-1}, \dots, u_1) & := & Q_{X_i|X_{i-1},\dots,X_1}(u_i|x_{i-1}, \dots, x_1) =: x_i, \\ & & 0 < u_i < 1, \quad 2 \leq i \leq d. \end{array}$$

  the successive conditional q.f. of the conditional distributions of X_i given $(X_{i-1}, \dots, X_1)$, for $1 \leq i \leq d$, see Rüschendorf [42]. Then, letting

$$\tilde{\mathbf{X}} := \mathbf{Q}_{\mathbf{U}\to\mathbf{X}}(\mathbf{U}) = (Q_1(U_1), \dots, Q_d(U_d|U_{d-1}, \dots, U_1)),$$

  one obtains a random vector $\tilde{\mathbf{X}}$ equal to $\mathbf{X}$ in distribution,

$$\tilde{\mathbf{X}} \stackrel{d}{=} \mathbf{X},$$

  i.e. the multivariate analogue of (8).

- Starting from a vector $\mathbf{X} = (X_1, \dots, X_d)$ with prescribed c.d.f. F, the reciprocal transformation (known as *Rosenblatt's transform* in the continuous case [36] and generalised to the general case in Rüschendorf [42]),

$$\begin{array}{rcl} \mathbf{R}_{\mathbf{X}\to\mathbf{U}}(., \mathbf{V}) : \mathbb{R}^d & \to & [0,1]^d \\ \mathbf{X} & \mapsto & \mathbf{R}_{\mathbf{X}\to\mathbf{U}}(\mathbf{X}, \mathbf{V}) \end{array}$$

is defined similarly as in (10) via the set of successive extended conditional c.d.f.

$$(F_{X_i|X_{i-1},\dots,X_1}(x_i,\lambda_i|x_{i-1},\dots,x_1), 1 \le i \le d), \quad 0 < \lambda_i < 1$$

and an additional *randomizer* vector $\mathbf{V} = (V_1, \dots V_d)$, made of i.i.d. marginals $U_{[0,1]}$ r.v., also jointly independent of $\mathbf{X}$. The *multivariate conditional distributional transform* is the randomized transformation

$$\mathbf{R}_{\mathbf{X}\to\mathbf{U}}(\mathbf{X},\mathbf{V}) := (F_{X_1}(X_1,V_1),\dots,F_{X_d|X_{d-1},\dots,X_1}(X_d,V_d|X_{d-1},\dots,X_1)).$$

Then, one has the analogue of (10): if

$$\mathbf{U} := \mathbf{R}_{\mathbf{X}\to\mathbf{U}}(\mathbf{X},\mathbf{V}), \tag{12}$$

then $\mathbf{U}$ is uniform on the unit cube and

$$\mathbf{Q}_{\mathbf{U}\to\mathbf{X}}(\mathbf{U}) = \mathbf{X} \quad \text{a.s.}$$

Again, this view encapsulates Laplace's view on randomness and is similar to the engineers' approach on modeling time series, (see Priestley [27], chapter 2): the most unpredictable time series is a strong white noise, viz. a sequence (U_i) of i.i.d. r.v. with a common prescribed distribution (here uniform on $[0,1]$, but which is often taken standard Gaussian in the context of time series). Hence, starting from such a sequence (U_i) of i.i.d. r.v. considered as a "source of randomness", Nature generates successively the next output X_{i+1} from the "past" realizations $(X_1,\dots,X_i)$ by a random mechanism involving an independent U_{i+1}. Such random mechanism is described by the "response functions" formed by the successive conditional q.f. One obtains a "Markov (quantile) regression representation" of $\mathbf{X} \sim F$ from the source of i.i.d U_i r.v. Reciprocally, one can consider a stochastic temporal model, made of the successive extended conditional c.d.f., has captured all the stochastic dependence in a vector $\mathbf{X}$, if it can transform the latter vector into strong white noise, i.e. into a sequence of i.i.d. r.v. with a prescribed univariate reference distribution(here uniform).

2.2.2 Copula representations

In an approach dual to the multivariate conditional quantile representation of the previous section 2.2.1, one may start from a vector $\mathbf{X} = (X_1,\dots,X_d)$ with given multivariate c.d.f. F, and wish to obtain a vector $\mathbf{U} = (U_1,\dots,U_d)$, which is no longer made of independent marginals as in (12), but captures the "dependence", irrespectively of the marginals. This is obtained by standardizing the marginals of $\mathbf{X}$ by using the *marginal* distributional transforms, thus obtaining a vector $\mathbf{U}$, whose c.d.f. is a *copula* function, as is explained now.

- A primer on copulas as probabilistic transforms:

For $\mathbf{X} \sim F$, denote its corresponding vector of marginal cdfs by $\mathbf{G} = (G_1, \dots, G_d)$, namely

$$G_i(x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty).$$

Recall that a d-dimensional copula function $C : [0,1]^d \mapsto [0,1]$ is defined analytically as a grounded, $d-$increasing function, with uniform marginals whose domain is $[0,1]^d$ (see Nelsen [23]). Alternatively, it can be defined probabilistically as the restriction to $[0,1]^d$ of the multivariate cdf of a random vector $\mathbf{U}$, called a *copula representer*, whose marginals are uniformly distributed on $[0,1]$ (see Rüschendorf [42, 43]). Their interest stems from Sklar's theorem (see [52, 53]), which asserts that, for every random vector $\mathbf{X} \sim F$, there exists a copula function *connecting, or associated with* $\mathbf{X}$, in the sense that:

Theorem 2.1. *For every multivariate cdf F, with marginal cdfs $\mathbf{G}$, there exists some copula function C such that*

$$F(\mathbf{x}) = C(\mathbf{G}(\mathbf{x})), \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{13}$$

Conversely, if C is a copula function and $\mathbf{G} = (G_1, \dots, G_d)$ a vector of marginal univariate distribution functions, then the function F defined by (13) is a joint distribution function with marginals $\mathbf{G}$.

When $\mathbf{G}$ is continuous, the copula C associated with $\mathbf{X}$ in relation (13) is unique and can be defined from F either analytically by $C = F \circ \mathbf{G^{-1}}$, where $\mathbf{G^{-1}} = (G_1^{-1}, \dots, G_d^{-1})$ is the vector of marginal quantile functions, or probabilistically as the cdf of the multivariate marginal probability integral transforms, namely $C(\mathbf{u}) = P(\mathbf{G}(\mathbf{X}) \leq \mathbf{u})$, $\mathbf{u} \in [0,1]^d$. Whenever discontinuity is present, C is no longer unique: in other words C, as a functional parameter, is not identifiable from the multivariate cdf F alone. In such a case, the most natural way to derive a probabilistic construction of a copula representer $\mathbf{U}$ associated with $\mathbf{X}$ is to use the the d-variate *marginal* distributional transform: set

$$\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$$

where is the vector of extended marginal cdfs, and $\mathbf{V}$ is a vector of uniform $[0,1]$ marginals (i.e. its cdf is itself a copula function), independent of $\mathbf{X}$. Then, the cdf C of $\mathbf{U}$ is a copula function which satisfies (13), see Moore and Spruill [21], Rüschendorf [38, 42, 43], Faugeras [11, 12]. The distribution function of $\mathbf{V} = (V_1, \dots, V_d)$ can be any copula, but the most natural choice is to choose the independent one, so that dependence measures computed on $\mathbf{U}$ matches those computed on $\mathbf{X}$, see [12]. Hence, one can view again the pair $(\mathbf{G^{-1}}(.), \mathbf{G}(., \mathbf{V}))$ as reciprocal transformations between $\mathbf{X}$ and its copula representer $\mathbf{U}$,

$$\begin{array}{ccc} \mathbf{U} & \xrightarrow{\mathbf{G^{-1}}} & \mathbf{X} \\ \mathbf{U} & \xleftarrow{\mathbf{G}(.,\mathbf{V})} & \mathbf{X}. \end{array} \tag{14}$$

- Empirical copulas:

  If F is unknown, but one has instead a sample $\mathbf{X}_1, \mathbf{X}_2, \ldots$ of copies distributed according to F on a probability space $(\Omega, \mathcal{A}, P)$, one can define the ecdf F_n,

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} 1_{\mathbf{X_i} \leq \mathbf{x}},$$

  and the corresponding vector of marginal ecdfs $\mathbf{G}_n$. Sklar's theorem therefore entails that there exists some copula function C_n associated with F_n. As the ecdf is discrete, C_n is no longer unique and can no longer be defined, in parallel with the continuous case, as $C_n^* := F_n \circ \mathbf{G}_n^{-1}$, or as $C_n^{**}(\mathbf{u}) := P^*(\mathbf{G}_n(\mathbf{X}_n^*) \leq \mathbf{u})$, with $\mathbf{X}_n^* \sim F_n$, conditionally on the sample, and where P^* is the corresponding conditional probability (more on this below). Indeed, C_n^* and C_n^{**} do not have uniform marginals and hence are not genuine copula functions associated with F_n. C_n^* and C_n^{**} are versions of the improperly called empirical "copula" functions, introduced by Rüschendorf [37] under the name of multivariate rank order function and Deheuvels [9, 10] under the name of empirical dependence function.

  When F is continuous, the disadvantage of estimating $C = F \circ \mathbf{G^{-1}}$ by estimators which are not proper, in the sense that they do not belong to the same functional class of the parameter to be estimated, is mitigated by the fact that these estimators coincide, with any copula function associated with F_n on the grid of points $\mathbf{u}_k = (k_1/n, \ldots, k_d/n)$ for $k_1, \ldots, k_d = 0, \ldots, n$; see Deheuvels [10]. Moreover, any version of the corresponding empirical "copula" process weakly converges, see e.g. Fermanian et al. [13], Deheuvels [10], or Rüschendorf [37]. Hence, in the continuous case, the choice of which "empirical copula" function to use is often of little relevance for statistical purposes.

  However, we will see in section 4 that defining the empirical copula as a genuine copula function will be a key element in the construction of empirical quantile morphisms, even in the continuous case. Hence, let us define the empirical copula representer, conditionally on the sample, as follows: on an extra probability space $(\Omega^*, \mathcal{A}^*, P^*)$, let $\mathbf{X}_n^* \sim F_n$. Set

$$\mathbf{U}_n := \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}), \tag{15}$$

  the multivariate distributional transforms for the ecdf F_n, with independent randomisation $\mathbf{V}$. Denote as C_n the cdf of $\mathbf{U}_n$, i.e. the copula function associated with F_n.

3 Quantile as Markov morphisms: an algebraic categorical view

What often matters in probability and statistics is not the random elements $\mathbf{X}$ *per se*, but the distribution they induce $P^{\mathbf{X}}$ and the properties of the latter.

Hence, it is advantageous to see the transformations (14), (12), (11) of random elements of section 2 as *transformations of probability measures*,

$$P^{\mathbf{X}} \leftrightarrows P^{\mathbf{U}}.$$

(Such view can be rendered rigorous, subject to some measure theoretic subtleties which we leave aside, see details in Zolotarev [62] chapter 1, Rachev [30], chapter 2 or the introduction in Cencov [4]). This final change of perspective will allow us to eventually motivate the forthcoming definition of the quantile object from an abstract algebraic viewpoint, inspired by category theory.

3.1 The category of Markov morphisms

Informally, category theory is made of "objects" or abstract sets denoted by $A, B, C, \dots$ and a system of mappings, morphisms or "arrows" $\alpha, \beta, \dots$ of the objects into one another. It will often be necessary to precise the domain and codomain of morphisms by subscripts, so that $A \xrightarrow{\alpha} B$, be denoted by α_{AB}. Such system of morphisms must obey the following two axioms:

- Associativity of composition: morphisms with compatible domains and codomains can be composed, i.e. if $\alpha_{AB} : A \to B$, $\beta_{BC} : B \to C$ then $\alpha_{AB} \circ \beta_{BC} : A \to C$ is also a morphism. (We will use Cencov's [4] left-to right convention so that $A\alpha = B$ (and *not* $\alpha(A) = B$) denotes the "transformation" of A into B under the "action" of α, so that composition of morphisms is reversed from the usual $\circ$ composition operation, as in e.g. Lawvere [19]). Moreover, the composition law is associative, i.e. if $A \xrightarrow{\alpha} B \xrightarrow{\beta} C \xrightarrow{\gamma} D$, then $(\alpha_{AB} \circ \beta_{BC}) \circ \gamma_{CD} = \alpha_{AB} \circ (\beta_{BC} \circ \gamma_{CD})$.
- Identity law: For every object A, the system includes the identical mapping $id_A = \epsilon_{AA}$ which is both the right and left identity: $\epsilon_{AA} \circ \alpha_{AB} = \alpha_{AB}$, and $\beta_{BA} \circ \epsilon_{AA} = \beta_{BA}$.

The system of morphisms forms an *abstract category* and the pair (objects, morphisms) is called a *concrete category*: it is the conceptual framework to carry the idea that a class of objects or "figure" A can be transformed into other another figure B via "motions" α. It paves the way to study the properties of those figures which remain invariants under structure-preserving transformations, see [4, 18, 19, 17] for readable introductions.

In his seminal book, Cencov [4] show how statistical inference can be studied from such a viewpoint. To that purpose, let us recall the definition of a *Markov probability kernel* or Markov morphism:

Definition 3.1. *Let $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces, a function $\mathcal{K}_{12} : \Omega_1 \times \mathcal{A}_2 \mapsto [0, 1]$ is a Markov probability kernel (or transition probability distribution) from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ iff*

i) for every $\omega_1 \in \Omega_1$, $A_2 \in \mathcal{A}_2 \to \mathcal{K}_{12}(\omega_1, A_2)$ is a probability measure on $\mathcal{A}_2$;

ii) for every $A_2 \in \mathcal{A}_2$, $\omega_1 \in \Omega_1 \to \mathcal{K}_{12}(\omega_1, A_2)$ is a $\mathcal{A}_1$-measurable function.

Markov kernels $\mathcal{K}_{12}$ will also be denoted *Markov morphisms*: $\mathcal{K}_{12}$ induces two morphisms preserving the algebraic structure (i.e. homomorphisms), which will be denoted by the same letter $\mathcal{K}_{12}$:

- a positive bounded linear operator on the convex set of probability measures : $\mathcal{K}_{12}$ transforms a probability measure P_1 on $(\Omega_1, \mathcal{A}_1)$ into a probability measure P_2 on $(\Omega_2, \mathcal{A}_2)$, by acting on the right on measures as

$$P_2(.) := (P_1\mathcal{K}_{12})(.) := \int_{\Omega_1} P_1(d\omega_1)\mathcal{K}_{12}(\omega_1, .), \tag{16}$$

  see Cencov [4] lemma 5.2 p. 67. Symbolically,

$$P_1 \xrightarrow{\mathcal{K}_{12}} P_2.$$

- a positive bounded linear operator $\mathcal{K}_{12}$ on the vector space $\mathcal{F}_b(\Omega_2, \mathcal{A}_2)$ of bounded measurable functions $f_2 : (\Omega_2, \mathcal{A}_2) \mapsto \mathbb{R}$ into the vector space of bounded measurable functions $\mathcal{F}_b(\Omega_1, \mathcal{A}_1) := \{f_1 : (\Omega_1, \mathcal{A}_1) \mapsto \mathbb{R}\}$, by acting on the left on functions as

$$f_1(.) := (\mathcal{K}_{12}f_2)(.) := \int_{\Omega_2} \mathcal{K}_{12}(., d\omega_2)f_2(\omega_2), \tag{17}$$

  see [4] lemma 5.1 p. 66. Symbolically,

$$f_1 \xleftarrow{\mathcal{K}_{12}} f_2.$$

- and one has commutation, see [4] lemma 5.3 p. 68, i.e.

$$(P_1\mathcal{K}_{12})f_2 = P_1(\mathcal{K}_{12}f_2). \tag{18}$$

  where

$$P(f) := \int f dP := \langle P, f\rangle \tag{19}$$

  is the expectation of f w.r.t. P, viz. the duality bracket between measures and functions. In view of (16, 17, 18), the action of $\mathcal{K}_{12}$ on measures and functions will be written without parentheses nor brackets in the remainder.

Such Markov probability kernels $\mathcal{K}_{12} : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ and $\mathcal{K}_{23} : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ obey the composition law,

$$\mathcal{K}_{13}(\omega_1, A_3) := \int_{\Omega_2} \mathcal{K}_{12}(\omega_1, d\omega_2)\mathcal{K}_{23}(\omega_2, A_3) \tag{20}$$

which is associative ([4] lemma 5.4 and 5.6), and the Dirac kernel $\mathcal{I} : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ defined by

$$\mathcal{I}(\omega, A) = \delta_\omega(A), \tag{21}$$

corresponding to the Dirac measure in ω, is the identity on $(\Omega, \mathcal{A})$ for the composition law (20), ([4] lemma 5.8). In other words, the system of Markov morphisms is an *abstract category* ([4] theorem 5.1) and together with the class of probability (or signed) measures a *concrete category* ([4] theorem 5.2 and lemma 5.9). In view of (16, 17, 18, 20), we drop the composition symbol $\circ$ and denote composition by mere juxtaposition, viz. $\mathcal{K}_{13} = \mathcal{K}_{12}\mathcal{K}_{23}$. Let us state out this key result as a theorem:

Theorem 3.2 (Cencov's theorem 5.2)**.** *The class of objects $\mathcal{P}(\Omega, \mathcal{A})$ of probability measures on $(\Omega, \mathcal{A})$ with the system of Markov morphisms $\{\mathcal{K}\}$ of Markov probability kernels forms the concrete category CAP of all probability measures.*

Remark 1. *By duality (18), (19), the class of objects $\mathcal{F}_b(\Omega, \mathcal{A})$ of bounded, measurable, real-valued functions $f : \Omega \to \mathbb{R}$ with the system of Markov morphisms $\{\mathcal{K}\}$ of Markov probability kernels forms the concrete category CAF of all bounded measurable functions.*

3.2 Quantiles as Markov morphisms: qualitative aspects

The conceptual framework introduced in the previous subsection allows us to recast the problem of defining a quantile object of section 1 and to review the transformations between random variables of section 2 at the right categorical level of Markov morphisms.

Indeed, such Markov morphisms allow to subsume probability measures and (the measure induced by) random variables into the same abstract conceptual object:

- a probability measure P_2 on $(\Omega_2, \mathcal{A}_2)$ is simply a constant Markov kernel $\mathcal{I}_P : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$, defined as,

$$\mathcal{I}_{P_2}(\omega_1, A_2) = P_2(A_2). \tag{22}$$

- a measurable function $f_{12} : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ can be described (embedded) as a degenerate Markov morphism $\mathcal{I}^{f_{12}} : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$, as

$$\mathcal{I}^{f_{12}}(\omega_1, A_2) = \delta_{f_{12}(\omega_1)}(A_2) = \delta_{\omega_1}(f_{12}^{-1}(A_2)) \tag{23}$$

so that the image measure $P_2(.) := P_1 \circ f_{12}^{-1}(.)$ on $(\Omega_2, \mathcal{F}_2)$ induced by f_{12} from the measure P_1 on $(\Omega_1, \mathcal{F}_1)$ (or in the push-forward notation $f_{12}\#P_1 =: P_2$) simply writes as a composition (20) of Markov morphisms,

$$P_2 = P_1\mathcal{I}^{f_{12}} = \mathcal{I}_{P_1}\mathcal{I}^{f_{12}}.$$

For further reference, let us single out this family of degenerate Markov morphisms by stating out a definition:

Definition 3.3. *A Markov morphism $\mathcal{K}_{12}$ from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ is of degenerate type if there exists a measurable funtion $f_{12} : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ s.t.*

$$\mathcal{K}_{12} = \mathcal{I}^{f_{12}}.$$

Such notation is consistent with the identity morphism $\mathcal{I}$ on $(\Omega, \mathcal{A})$ of (21), as the Dirac kernel can be expressed as $\mathcal{I} = \mathcal{I}^{id}$, where $id : \Omega \to \Omega$ is the identity function.

As a consequence, the transformations between random vectors of section 2, can be reformulated as transformations between measures through a Markov morphism. In particular, we already noted that the univariate quantile transform (8), $Q_X : U \to X$ as a mapping between random variables, can be construed as a morphism $\mathcal{I}^{Q_X}$ between univariate measures,

$$P^U \xrightarrow{\mathcal{I}^{Q_X}} P^X,$$

where $\mathcal{I}^{Q_X}$ is a Markov morphism of the degenerate type (3.3), whereas its reciprocal, the distributional transform (10) $F(., V) : X \to V$ as a *randomized* transform between univariate random variables, can be construed as a genuine, non degenerate Markov morphism $\mathcal{D}_X$

$$P^U \xleftarrow{\mathcal{D}_X} P^X,$$

where $\mathcal{D}_X(x, A)$ is the conditional probability of $U := F(X, V) \in A$ given $X = x$. Let us define similarly $\mathcal{I}^{\mathbf{G}^{-1}}$ and $\mathcal{D}_{\mathbf{X}}$ their multivariate counterparts transforming corresponding to (14), i.e. transforming a multivariate $P^{\mathbf{X}}$ into its copula representer distribution $P^{\mathbf{U}} = P^{\mathbf{G}(X,V)}$,

$$\begin{array}{ccc} P^{\mathbf{U}} & \xrightarrow{\mathcal{I}^{\mathbf{G}^{-1}}} & P^{\mathbf{X}} \\ P^{\mathbf{U}} & \xleftarrow{\mathcal{D}_{\mathbf{X}}} & P^{\mathbf{X}} \end{array} \tag{24}$$

Remark 2 (Markov morphisms as degenerate Markov morphism on an enlarged probability space)**.** *Note that a randomised transform between random vectors (i.e. genuine Markov morphisms of the nondegenerate type) could also be written as a purely functional transform (i.e. as a degenerate Markov morphism), at the price of having to enlarge the probability space.*

For example, for the univariate distributional transform (10), enlarge $(\Omega, \mathcal{A}, P)$ *to* $(\Omega \times [0,1], \mathcal{A} \otimes \mathcal{B}([0,1]), P \otimes \lambda)$*, denote* F_X *the c.d.f. of* X *and* $\lambda = P^V$ *the Lebesgue measure on* $[0,1]$*, transfer all previously defined random elements on this new, enlarged probability space, and consider the bivariate mapping*

$$\begin{aligned} F_X(.,.) : \mathbb{R} \times [0,1] &\to [0,1] \\ (X, V) &\mapsto F_X(X, V). \end{aligned}$$

and the corresponding bivariate product mapping $F_X \otimes id$*,*

$$\begin{aligned} F_X \otimes id : \mathbb{R} \times [0,1] &\to [0,1] \times [0,1] \\ (X, V) &\mapsto (F_X(X, V), V). \end{aligned}$$

Then, with $U := F_X(X, V)$ *and* $\pi : (U, V) \mapsto U$ *the projection mapping on the first coordinate,*

$$P^X \otimes \lambda \xrightarrow{\mathcal{I}^{F_X \otimes id}} P^{(U,V)} \xrightarrow{\mathcal{I}^{\pi}} P^U = \lambda.$$

Hence, the univariate distributional transform (10) can be construed as the composition of two degenerate Markov morphisms, i.e. as a degenerate Markov morphism, on the enlarged space.

More generally, see Thorisson [56] chapter 3, Kallenberg [16] lemma 2.22, lemma 5.9 and theorem 5.10 p. 89 and [50] lemma 1.1 p. 18 and lemma 1.3 p. 20 for a rigorous formulation of the principle that "a randomized decision for an experiment $\mathcal{E}$ is a non-randomized decision, but for an experiment $\tilde{\mathcal{E}}$ which is an "extension" of $\mathcal{E}$" ([50] p. 17). The same remark apply to the multivariate conditional quantile (11), distributional (12), and copula (14) transforms of section 2. However, from the categorical perspective, limiting oneself to purely functional transforms is awkward and we prefer to unify all kind of transformations into the same category of Markov morphism.

Eventually, note also that when we will represent the Markov morphism corresponding to the transformations (14) and (15) by the random elements $\mathbf{U} = \mathbf{G}(\mathbf{X}, \mathbf{V})$ and $\mathbf{U}_n = \mathbf{G_n}(\mathbf{X_n^*}, \mathbf{V})$, we use the same random vector $\mathbf{V}$ in both cases, in order to obtain a.s. convergence (see the forthcoming theorem 5.1 in section 5).

We are now in a position to formulate a (preliminary) definition of the object multivariate quantile as a Markov morphism, which combines the idea (2) in section 1 of quantile as an inverse, and those of section 2 of quantile as a reciprocal (randomized) transformation between vectors (or measures), one of which being thought of as a reference distribution (see (8), (10) in section 2):

Definition 3.4 (Preliminary). *Let $(\mathcal{S}, \mathcal{B}(S))$ and $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be two measurable spaces. Consider as object on $\mathcal{S}$, the one set consisting of a single probability measure $1 := \{P^{\mathbf{S}}\}$, which is thought as a reference distribution, and as objects on $\mathcal{X}$ the collection $\mathcal{P}(\mathcal{X}, \mathcal{A}) := \{P^{\mathbf{X}}\}$ of all probability measures on $\mathcal{X}$. A Quantile morphism $\mathcal{Q}$ of $P^{\mathbf{X}}$ w.r.t. $P^{\mathbf{S}}$ is an isomorphism*

$$P^{\mathbf{S}} \xrightarrow{\mathcal{Q}} P^{\mathbf{X}},$$

whose inverse $\mathcal{R}$,

$$P^{\mathbf{S}} \xleftarrow{\mathcal{R}} P^{\mathbf{X}}$$

will be called a Rank morphism of $P^{\mathbf{X}}$ w.r.t. $P^{\mathbf{S}}$. In other words, the pair of Markov morphisms $(\mathcal{Q}, \mathcal{R})$ satisfy $\mathcal{QR} = \mathcal{I}_{\mathbf{S}}$, and $\mathcal{RQ} = \mathcal{I}_{\mathbf{X}}$, where $\mathcal{I}_{\mathbf{S}}, \mathcal{I}_{\mathbf{X}}$ are the identity (21) on $\mathcal{S}, \mathcal{X}$, respectively.

Remark 3. *We used the term rank morphism instead of distributional (or probability integral, or c.d.f) morphism to agree with the terminology of [5], see section 4 below.*

We temporarily leave aside the question of existence and unicity of these Quantile and Rank morphisms, since this qualitative view of reciprocal transformations of measures will be turned into a *quantitative* problem, via mass transportation theory, as we now show.

3.3 Quantitative transformation of measures via Mass Transportation

The *Monge-Kantorovich optimal transportation problem* aims at finding a joint measure $P^{\mathbf{X},\mathbf{Y}}$ on the product measurable space, say $(\mathcal{X}\times\mathcal{Y}=\mathbb{R}^d\times\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\otimes\mathcal{B}(\mathbb{R}^d))$, with prescribed marginals $(P^{\mathbf{X}}, P^{\mathbf{Y}})$, which is the solution of the optimisation problem:

$$k_c(P^{\mathbf{X}}, P^{\mathbf{Y}}) := \inf_{P^{\mathbf{X},\mathbf{Y}}\in\mathcal{P}(P^{\mathbf{X}},P^{\mathbf{Y}})} P^{\mathbf{X},\mathbf{Y}}[c(\mathbf{X},\mathbf{Y})], \tag{25}$$

where $c : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^+$ is a cost function and the infimum is on the set $\mathcal{P}(P^{\mathbf{X}}, P^{\mathbf{Y}})$ of joint distribution with marginals $P^{\mathbf{X}}$, $P^{\mathbf{Y}}$. Informally, mass at $\mathbf{x}$ of $P^{\mathbf{X}}$ is transported to $\mathbf{y}$, according to the conditional distribution $P(d\mathbf{y}|\mathbf{x})$ of the *transportation plan* $P^{\mathbf{X},\mathbf{Y}} \in \mathcal{P}(P^{\mathbf{X}}, P^{\mathbf{Y}})$, in order to recover $P^{\mathbf{Y}}$ while minimising the average cost of transportation $P^{\mathbf{X},\mathbf{Y}}[c(\mathbf{X},\mathbf{Y})]$. See Rachev and Ruschendorf [32], Villani [59, 60] for book-length treatment on the subject, Rachev [28], Rüschendorf [41, 40] for survey articles. This topic is closely related to coupling ([56]) and probability metrics ([62, 30, 31]).

The related *Monge transportation problem* is when one looks for a solution of (25) which is "deterministic" in the sense that the laws of $(\mathbf{X},\mathbf{Y})$ are restricted to those of $(\mathbf{X},\mathbf{H}(\mathbf{X}))$ for a measurable *transportation map* $\mathbf{H} := \mathbf{H}_{P^{\mathbf{X}}\to P^{\mathbf{Y}}} : \mathbb{R}^d \mapsto \mathbb{R}^d$, s.t. $P^{\mathbf{H}(\mathbf{X})} = P^{\mathbf{Y}}$, so that

$$m_c(P^{\mathbf{X}}, P^{\mathbf{Y}}) := \inf_{\mathbf{H}:P^{\mathbf{H}(\mathbf{X})}=P^{\mathbf{Y}}} P^{\mathbf{X}}[c(\mathbf{X},\mathbf{H}(\mathbf{X}))]. \tag{26}$$

By disintegrating the transportation plan $P^{\mathbf{X},\mathbf{Y}}$ into the fibered product

$$P^{\mathbf{X},\mathbf{Y}} = P^{\mathbf{X}} \otimes P^{\mathbf{Y}}_{\mathbf{x}},$$

where $P^{\mathbf{Y}}_{\mathbf{x}}$ is a regular conditional distribution (i.e. a Markov kernel) of $\mathbf{Y}$ given $\mathbf{X}=\mathbf{x}$, it is clear that mass transportation can be translated in the language of Markov morphisms: the Monge-Kantorovich optimal transportation problem amounts to finding a Markov morphism $\mathcal{K} = P^{\mathbf{Y}}_{\mathbf{x}}$,

$$P^{\mathbf{X}} \xrightarrow{\mathcal{K}} P^{\mathbf{Y}},$$

which is a genuine Markov kernel, whereas Monge optimal transportation problem amounts to finding a degenerate Markov morphism $\mathcal{I}^{\mathbf{H}}$ of the kind (23) induced by a transportation map $\mathbf{H}$,

$$P^{\mathbf{X}} \xrightarrow{\mathcal{I}^{\mathbf{H}}} P^{\mathbf{Y}}.$$

The cost function c is often specialised to the squared euclidean distance, $c(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}||^2$, which yields the so-called L_2 *Wasserstein probability distance*, see e.g. [6]. The characterization of the optimal L_2 solution of (25) was given by theorem 1 in Rachev and R#uschendorf [44], and can be rewritten in the language of Markov morphisms as follows: for $P^{\mathbf{X}}$ almost all $\mathbf{x}$, there exists some l.s.c. convex function $\psi : \mathbb{R}^d \to \mathbb{R}$, s. t.

$$supp(\mathcal{K}(\mathbf{x}, .)) \subset \partial\psi(\mathbf{x}),$$

where *supp* stands for the support and ∂ for the subgradient, see [44]. In particular, if $P^{\mathbf{X}} << \lambda^d$, this result implies Brenier's theorem [34] on existence and unicity $P^{\mathbf{X}}$-a.e. of Monge's transportation map $\mathbf{H}$, while additional existence of second moments of both $P^{\mathbf{X}}$ and $P^{\mathbf{Y}}$ yields strict unicity. One then obtains, via Kantorovich-Fenchel-Legendre duality, the transportation maps $\mathbf{H}$ and its reciprocal transformation $\mathbf{H}^{\leftarrow}$ corresponding to

$$P^{\mathbf{X}} \overset{\mathcal{I}^{\mathbf{H}^{\leftarrow}}}{\longleftarrow} P^{\mathbf{Y}},$$

as gradients of ψ and its dual ψ^*,

$$\mathbf{H} = \nabla\psi, \quad \mathbf{H}^{\leftarrow} = \nabla\psi^* \tag{27}$$

see [44, 7, 28, 41, 39].

Remark 4. *Rüschendorf [39] shows how one can recover the multidimensional conditional quantile and distributional transform from such a mass transformation problem. Copulas, completely monotone dependent random variables and Fréchet-Hoeffding bounds also arise from mass transportation, see Rüschendorf [40, 43].*

3.4 Towards a structural point of view

One of the main interest of category theory is to propose a framework which allows to study which transformations of the objects in the category preserves the structures attached to these objects, i.e. which leave invariant the properties of the objects, see Cencov [4] paragraph 4 to 8.

As announced in section 1, we are now equipped with the right conceptual tools to substantiate our claim that the multivariate q.f. and related depth proposals of subsection 1.3 should be evaluated according to their *structural properties* instead of their method. To that purpose, let us reexamine the properties of q.f. listed in subsection 1.2, classify them according to this structural point of view, and translate them in categorical terms, so that we can let emerge a set of desirable axioms that multivariate quantile and rank objects should obey (some of which will be loosely stated).

- Algebraic Structure:
  - [A1] Identifiability / Isomorphism:
    in view of (4), (Q_X, F_X) characterizes their parent distribution. In the language of Markov morphisms, the pair of isomorphims $(\mathcal{Q}, \mathcal{R})$ in definition 3.4 are precisely achieve such characterization: $P^{\mathbf{S}}\mathcal{Q} = P^{\mathbf{X}}$, and $P^{\mathbf{S}}\mathcal{R} = P^{\mathbf{X}}$.

– [A2] Algebraic compatibility:

We viewed in section 1 that q.f. have good algebraic properties. Moreover, the collection of probability measures on some fixed measurable space forms a convex subset of the vector space of signed measures. Therefore, multivariate extensions of quantile objects should be compatible with addition and scalar multiplication in a way that reflect the underlying structure of the collection of measures. In view of (20), (16) and (17), and the fact that the set of Markov morphisms between two fixed measurable space is convex, such desideratum will be automatically satisfied if the objects quantile and rank are taken as Markov morphisms.

- Ordering structure:

– [O1] Galois connection between two ordered spaces.

The fundamental property of univariate q.f. (2) and (3), can be illuminated by introducing the notion of Galois connection (see Blyth [2], chapter 1 and also the related idea of residuated mapping):

Definition 3.5. *Let* $(X, \leq_X)$*,* $(Y, \leq_Y)$ *be two ordered sets and* $L : X \mapsto Y, U : Y \mapsto X$ *be a pair of mappings. Then* $(X, \leq_X), (Y, \leq_Y), L, U$ *is an isotone Galois connection*

* *iff for every* $x \in X$ *and every* $y \in Y$ $L(x) \leq_Y y \Leftrightarrow x \leq_X U(y)$
* *iff* L, U *are monotone (or isotone) and for every* $x \in X$ *and every* $y \in Y$*,* $x \leq_X U(L(x)) \Leftrightarrow L(U(y)) \leq_Y y$*.*

Indeed, consider the two ordered sets $(I, \leq_I)$, with $I = (0, 1)$ the unit interval and $\leq_I = \leq$ the usual order $\leq$ and $(\mathbb{R}, \leq_{\mathbb{R}})$ with its usual order $\leq_{\mathbb{R}} = \leq$. Then, for univariate q.f. properties (2) and (3), simply mean that $(I, \leq), (\mathbb{R}, \leq), Q_X, F_X$ is an isotone Galois connection. Therefore, the desirable property for Quantile and Rank morphisms: Quantile and Rank morphisms should form a Galois connection between two ordered spaces.

In the spirit of section 2, this concept of Galois connection can be reformulated in probabilistic terms: starting from a r.v. U uniformly distributed on $[0, 1]$, setting $X' := Q_X(U)$, one has that $U \in (0, F_X(x)]$ iff $X' \in]-\infty, x]$. In other words, starting from a reference distribution P^U, and a given a quantity of mass $0 \leq \tau \leq 1$, these considerations amount to construct on a common probability space some copies $X' \sim P^X = P^U \mathcal{I}^{Q_X}$ of X from a transformation of $U \sim P^U$ s.t. if $A_\tau :=]-\infty, x]$, is a subset of $\mathbb{R}$ with P^X mass τ, viz. $P^X(A_\tau) = \tau$, then $B_\tau = F_X(A_\tau)$ is a subset of $[0, 1]$ with P^U mass τ, with

$$Q_X(U) \in A_\tau \Leftrightarrow U \in F_X(A_\tau),$$

as

$$\begin{aligned}\tau &= P^X(A_\tau) = P(X' \in A_\tau) = P(Q_X(U) \in A_\tau)\\ &= P(U \in F_X(A_\tau)) = P(U \in B_\tau).\end{aligned}$$

In addition, the set of subsets $\mathcal{O}_I^+ := \{B_\tau, \tau \in [0,1]\}$, $\mathcal{O}_\mathbb{R}^+ := \{A_\tau, \tau \in [0,1]\}$ should have an order structure (for the inclusion) compatible with the order relations $\prec_I, \prec_\mathbb{R}$, in the sense that $\tau_1 \prec_I \tau_2 \Leftrightarrow A_{\tau_1} \subset A_{\tau_2} \Leftrightarrow B_{\tau_1} \subset B_{\tau_2}$.

Turning to the general case, the setting can be (loosely) formulated in terms of the Markov morphisms $(\mathcal{Q}, \mathcal{R})$ as follows: let $\mathcal{F}^*(\mathcal{S}) := \{\mathbf{s} \to 1_{B_\tau}(\mathbf{s})\}$ a collection of indicator function $\mathcal{S} \to \mathbb{R}$, indexed by $0 \leq \tau \leq 1$, where $\{B_\tau\} \subset \mathcal{B}(\mathcal{S})$ is a collection of measurable depth regions of $P^\mathbf{S}$ mass τ, $P^\mathbf{S}(B_\tau) = \tau$. Define $\mathcal{F}^*(\mathcal{X}) := \mathcal{R}(\mathcal{F}^*(\mathcal{S})) = \{A_\tau := \mathcal{R}1_{B_\tau}\}$, the image of $\mathcal{F}^*(\mathcal{S})$ by the Rank Morphism $\mathcal{R}$. Then,

$$P^\mathbf{X}(A_\tau) = P^\mathbf{X}(\mathcal{R}1_{B_\tau}) = (P^\mathbf{X}\mathcal{R})(1_{B_\tau}) = P^\mathbf{S}(B_\tau) = \tau,$$

i.e. A_τ is of $P^\mathbf{X}$ mass τ. These are depth "regions" (functions) for $\mathcal{X}$, see section 3.5 for a detailed description of the order structure, its preservation by Markov morphisms and its interpretation.

- [O2] Equivariance w.r.t. left-continuous univariate monotone transformation:

  in view of (6), one should have some form of scale invariance w.r.t. to a monotone non-decreasing univariate transformation. In the multivariate case, let $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^d$ be made of d univariate monotone non-decreasing functions $g_i : \mathbb{R} \to \mathbb{R}$, $1 \leq i \leq d$, viz. $\mathbf{g}(\mathbf{x}) = (g_1(x_1), \ldots, g_d(x_d))$. In the language of Markov morphisms, if $\mathcal{Q}$ and $\mathcal{Q}_\mathbf{g}$ are quantile morphisms of $\mathbf{X}$, $\mathbf{g}(\mathbf{X})$ respectively, i.e.

  $$\begin{aligned}P^\mathbf{S} &\xrightarrow{\mathcal{Q}} P^\mathbf{X} \xrightarrow{\mathcal{I}^\mathbf{g}} P^{\mathbf{g}(X)}\\ P^\mathbf{S} &\xrightarrow{\mathcal{Q}_\mathbf{g}} P^{\mathbf{g}(X)},\end{aligned}$$

  one should have commutativity of the composition diagram,

  $$\mathcal{Q}_\mathbf{g} = \mathcal{Q}\mathcal{I}^\mathbf{g}.$$

- Topological structure:

  - [T1] Compatibility with weak convergence of measures.

In view of (5), it is desirable that quantile generalisations should be compatible with some notion of weak convergence on the space of probability measures: Let $P^{\mathbf{X}_n}, P^{\mathbf{S}_n}$ a sequence of probability measures on $\mathcal{X}, \mathcal{S}$ respectively. Let $(\mathcal{Q}_n, \mathcal{R}_n)$, respectively $(\mathcal{Q}, \mathcal{R})$ their corresponding quantile and rank morphisms of definition 3.4,

$$\begin{aligned}P^{\mathbf{S}_n}\mathcal{Q}_n = P^{\mathbf{X}_n}, &\qquad P^{\mathbf{X}_n}\mathcal{R}_n = P^{\mathbf{S}_n},\\ P^\mathbf{S}\mathcal{Q} = P^\mathbf{X}, &\qquad P^\mathbf{X}\mathcal{R} = P^\mathbf{S}.\end{aligned}$$

One requires that $P^{\mathbf{X}_n} \xrightarrow{d} P^{\mathbf{X}} \Leftrightarrow P^{\mathbf{S}_n} \xrightarrow{d} P^{\mathbf{S}}$.

Combining the preliminary definition (3.4) and these desirable properties, we can state an eventual possible definition of quantile and rank morphisms:

Definition 3.6. *Let $P^{\mathbf{S}}$ be a fixed reference distribution and $P^{\mathbf{X}}$ be the distribution considered on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. A Quantile and Rank morphism of $P^{\mathbf{X}}$ w.r.t. $P^{\mathbf{S}}$ is a pair of Markov isomorphisms $(\mathcal{Q}, \mathcal{R})$*

$$\begin{array}{ccc} P^{\mathbf{S}} & \xrightarrow{\mathcal{Q}} & P^{\mathbf{X}} \\ P^{\mathbf{S}} & \xleftarrow{\mathcal{R}} & P^{\mathbf{X}}, \end{array}$$

compatible with the algebraic, topological and ordering structures given by the axioms [A1,A2,O1,O2,T1].

3.5 Order structure and depth regions

The concept of depth requires a way to quantify a measure of "remoteness" of a distribution from a "deepest" or most central point. The minimal requirements to translate these phenomenological notions into a mathematical concept seems to give a preorder structure (a reflexive, transitive relation $\leqslant_{\mathcal{S}}$) on the reference space $\mathcal{S}$. In order to be able to define such a "median" or "deepest point", it is necessary that the preordered space $(\mathcal{S}, \leqslant_{\mathcal{S}})$ possess a smallest element $\mathbf{0}$, so that one can evaluate the degree of "remoteness" of two points in $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$ w.r.t. $\mathbf{0}$. In other words, one consider that $(\mathcal{S}, \leqslant_{\mathcal{S}})$ is a preordered set with a universal lower bound $\mathbf{0}$, so that $(\mathcal{S} - \mathbf{0}, \leqslant_{\mathcal{S}})$ has the structure of a *downward directed set*, viz a preordered set s.t. every pair of elements has a lower bound. (See [8], [45] for background on order).

If $\mathcal{S}$ is chosen as a subset of $\mathbb{R}^d$ and has the algebraic and metric structure given by the usual Euclidean distance $||.||$, such framework is obtained by setting

$$\mathbf{s_1} \leqslant_{\mathcal{S}} \mathbf{s_2} \Leftrightarrow ||\mathbf{s}_1 - \mathbf{0}|| \leq ||\mathbf{s}_2 - \mathbf{0}|| \Leftrightarrow ||\mathbf{s}_1|| \leq ||\mathbf{s}_2||.$$

As in [5], let us choose $\mathcal{S}$ to be the unit ball of $\mathbb{R}^d$ of center $\mathbf{0}$. (Notice that $(\mathcal{S}, \leqslant_{\mathcal{S}})$ is not a lattice (two points on a same sphere have same radius and can not be distinguished) and also that this preorder is not compatible with the vector space structure of $\mathcal{S}$, in the sense that $\mathbf{s_1} \leqslant_{\mathcal{S}} \mathbf{s_2} \not\Rightarrow \mathbf{s_1} + \mathbf{s_3} \leqslant_{\mathcal{S}} \mathbf{s_2} + \mathbf{s_3}$ for $\mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3} \in \mathcal{S}$).

The general idea is as follows:

1. transfer this preorder on $\mathcal{S}$ into a stochastic preorder on $\mathcal{P}(\mathcal{S})$, the set of probability measures on $\mathcal{S}$, via an order embedding $\mathbf{s} \to \delta_{\mathbf{s}}$;

2. then use Markov kernels $(\mathcal{Q}, \mathcal{R})$ in Galois connection to transfer this preorder onto $\mathcal{P}(X)$;

3. obtains a preorder on $\mathcal{X}$ via the mapping $\delta_{\mathbf{x}} \to \mathbf{x}$, or at least a depth function or area on $\mathcal{X}$.

More precisely,

1. Step one:

    - Preorder interval: On $(\mathcal{S}, \leqslant_{\mathcal{S}})$, define the (pre-)order interval $[\mathbf{r}, \mathbf{t}] := \{\mathbf{s} \in \mathcal{S}, \mathbf{r} \leqslant_{\mathcal{S}} \mathbf{s} \leqslant_{\mathcal{S}} \mathbf{t}\}$, so that the closed balls $B_\tau := \{\mathbf{s} \in \mathcal{S}, |\mathbf{s}| \leq \tau\} = [\mathbf{0}, \mathbf{s}]$ for some $\mathbf{s} \in \mathcal{S}$ s.t. $|\mathbf{s}| = \tau$.
    - Down-sets : Recall that for $A \subset S$, A is a *down-set* if

    $$\mathbf{t} \in A, \mathbf{s} \leqslant_{\mathcal{S}} \mathbf{t} \Rightarrow \mathbf{s} \in A$$

    Denote the set of down-sets of $\mathcal{S}$ as $\mathcal{O}(\mathcal{S})$. Define the down-closure or order ideal of $A \subset S$ as the smallest down-set containing A: $A^{\downarrow} := \{\mathbf{s}, \mathbf{s} \leqslant_{\mathcal{S}} \mathbf{t} \text{ for some } \mathbf{t} \in A\}$. Hence, the ball $\mathbf{s}^{\downarrow} := \{\mathbf{s}\}^{\downarrow} = [\mathbf{0}, \mathbf{s}]$ is the *principal ideal* generated by $\mathbf{s}$. Set $\mathcal{O}^*(\mathcal{S}) = \{\mathbf{s}^{\downarrow}, \mathbf{x} \in \mathcal{S}\}$ the set of balls/principal ideals of $\mathcal{S}$.
    - Partial embedding into the powerset: Although $(\mathcal{S}, \leqslant_{\mathcal{S}})$ is only a preordered space, its set of principal down-sets $\mathcal{O}^*(\mathcal{S})$ is a partially ordered set (poset) included in the powerset $(2^{\mathcal{S}}, \subset)$, via the mapping

    $$\begin{array}{rcl} (\mathcal{S}, \leqslant_{\mathcal{S}}) & \to & (\mathcal{O}^*(\mathcal{S}) \subset 2^{\mathcal{S}}, \subset) \\ \mathbf{s} & \mapsto & \mathbf{s}^{\downarrow} = [\mathbf{0}, \mathbf{s}] \end{array}$$

    since

    $$\mathbf{s} \leqslant_{\mathcal{S}} \mathbf{t} \Leftrightarrow [\mathbf{0}, \mathbf{s}] \subset [\mathbf{0}, \mathbf{t}].$$

    (Notice however that $\mathbf{s} \mapsto [\mathbf{0}, \mathbf{s}]$ is not injective, hence one has only a partial embedding)
    Each element of $\mathcal{O}^*(\mathcal{S})$ is obviously measurable, so the embedding is in reality in $\mathcal{B}(\mathcal{S})$. This principal ideal $\mathcal{O}^*(\mathcal{S})$ makes up the collection of depth areas for $P^{\mathbf{S}}$: for each set/ball $B \in \mathcal{O}^*(\mathcal{S})$, there exists $\tau \in [0, 1]$ s.t. $P^{\mathbf{S}}(B) = \tau$. Hence, they capture the features required for their interpretation as central regions of $\mathcal{S}$ with a given $P^{\mathbf{S}}$ mass.
    - Order embedding on $\mathcal{P}(\mathcal{S})$: inspired from Massey [20], one can now define a stochastic preorder $\prec_{\mathcal{S}}$ on the space $\mathcal{P}(\mathcal{S})$ of probability measures compatible with the embedding $\mathbf{x} \to \delta_{\mathbf{x}}$, according to the following definition:

    Definition 3.7. *$\prec_{\mathcal{S}}$ is a stochastic preorder on $\mathcal{P}(\mathcal{S})$ if*

    i) $\prec_{\mathcal{S}}$ is a preorder on $\mathcal{P}(\mathcal{S})$;

    ii) $\mathbf{s} \to \delta_{\mathbf{s}}$ is an order-embedding: for all $\mathbf{s}, \mathbf{t} \in \mathcal{S}$, $\mathbf{s} \leqslant_{\mathcal{S}} \mathbf{t}$ holds iff $\delta_{\mathbf{s}} \prec_{\mathcal{S}} \delta_{\mathbf{t}}$.

    In our case, the stochastic preorder $\prec_{\mathcal{S}}$ on $\mathcal{P}(\mathcal{S})$ is defined via the previous embedding on the principal ideal $(\mathcal{S}, \leqslant_{\mathcal{S}}) \hookrightarrow (\mathcal{O}^*(\mathcal{S}), \subset)$:

    $$P^{\mathbf{S}_1} \prec_{\mathcal{S}} P^{\mathbf{S}_2} \Leftrightarrow P^{\mathbf{S}_1}([\mathbf{0}, \mathbf{s}]) \leq P^{\mathbf{S}_2}([\mathbf{0}, \mathbf{s}]), \quad \forall \mathbf{s} \in \mathcal{S}.$$

By duality or by considering the embedding $\mathcal{S} \hookrightarrow \mathcal{F}(\mathcal{S}) = \{f : \mathcal{S} \to \mathbb{R}\}$, obtained via the mapping $\mathbf{s} \to 1_{[\mathbf{0},\mathbf{s}]}$, the stochastic preorder $\prec_{\mathcal{S}}$ corresponds to an integral preorder with generator the set of indicator functions (see [22]): $\mathcal{F}^*(\mathcal{S}) := \{\mathbf{s} \to 1_{[\mathbf{0},\mathbf{s}]}\}$:

$$P^{\mathbf{S}_1} \prec_{\mathcal{S}} P^{\mathbf{S}_2} \Leftrightarrow P^{\mathbf{S}_1}(f) \leq P^{\mathbf{S}_2}(f), \quad \forall f \in \mathcal{F}^*(\mathcal{S}).$$

2. Step two: Galois connections for probability measures.

   Given a pair of Markov morphisms,

$$\begin{array}{ccc} P^{\mathbf{S}} & \xrightarrow{\mathcal{Q}} & P^{\mathbf{X}} \\ P^{\mathbf{S}} & \xleftarrow{\mathcal{R}} & P^{\mathbf{X}} \end{array}$$

   and a stochastic preorder $\prec_{\mathcal{S}}$ on $\mathcal{P}(\mathcal{S})$ defined by the function set $\mathcal{F}^*(\mathcal{S}) := \{\mathbf{s} \to 1_{[\mathbf{0},\mathbf{s}]}\}$ on $\mathcal{F}(\mathcal{S})$, the image of the latter by $\mathcal{R}$ defines a function set $\mathcal{F}^*(\mathcal{X}) := \mathcal{R}(\mathcal{F}^*(\mathcal{S})) = \{g := \mathcal{R}f, f \in \mathcal{F}^*(\mathcal{S})\}$ which in turns define an integral stochastic preorder $\prec_{\mathcal{X}}$ for $\mathcal{P}(\mathcal{X})$ so that

$$P^{\mathbf{S}}\mathcal{Q} \prec_{\mathcal{X}} P^{\mathbf{X}} \Leftrightarrow P^{\mathbf{S}} \prec_{\mathcal{S}} P^{\mathbf{X}}\mathcal{R}.$$

   Indeed,

$$\begin{aligned} P^{\mathbf{S}}\mathcal{Q} \prec_{\mathcal{X}} P^{\mathbf{X}} &\Leftrightarrow P^{\mathbf{S}}\mathcal{Q}(g) \leq P^{\mathbf{X}}(g), \forall g \in \mathcal{F}^*(\mathcal{X}) \\ &\Leftrightarrow P^{\mathbf{S}}\mathcal{Q}(\mathcal{R}f) \leq P^{\mathbf{X}}(\mathcal{R}f), \forall f \in \mathcal{F}^*(S) \\ &\Leftrightarrow P^{\mathbf{S}}(\mathcal{Q}\mathcal{R})f \leq P^{\mathbf{X}}\mathcal{R}(f), \forall f \in \mathcal{F}^*(\mathcal{X}) \\ &\Leftrightarrow P^{\mathbf{S}} \prec_{\mathcal{S}} P^{\mathbf{X}}\mathcal{R} \end{aligned}$$

   since $\mathcal{Q}R = \mathcal{I}_{\mathcal{S}}$. In other words, $(\mathcal{Q}, \mathcal{R})$ is a Galois connection between the preordered sets $(\mathcal{P}(\mathcal{S}), \prec_{\mathcal{S}})$ and $(\mathcal{P}(\mathcal{X}), \prec_{\mathcal{X}})$

3. Step three: depth areas in the $\mathcal{X}$ space.

   In turn, such preorder structure $(\mathcal{P}(\mathcal{X}), \prec_{\mathcal{X}})$ can sometimes be "descended" down to a preorder structure on $(\mathcal{X}, \leqslant_{\mathcal{X}})$, in case the mapping $\delta_{\mathbf{x}} \mapsto \mathbf{x}$ induces a compatible preorder structure. In particular, depth areas (set objects) in the $\mathcal{S}$ world, corresponding to the balls or principal down sets $\mathcal{O}^*(\mathcal{S})$ can become depths areas $\mathcal{R}(\mathcal{O}^*(\mathcal{S}))$ in the $\mathcal{X}$ world, in case the rank morphism $\mathcal{R}$ is of the degenerate type (3.3). Indeed, if $\mathcal{R} = \mathcal{I}^f$, for some $f : \mathcal{X} \to \mathcal{S}$ and $B_\tau \in \mathcal{O}^*(\mathcal{S})$ is of $P^{\mathbf{S}}$ mass τ, then

$$\mathcal{R}(B_\tau)(x) = 1_{f(x) \in B_\tau} = 1_{x \in f^{-1}(B_\tau)} = 1_{f^{-1}(B_\tau)}(x)$$

   is a function $\mathcal{X} \to \{0,1\} \cong 2$ isomorphic to the measurable set $f^{-1}(B_\tau)$ of $\mathcal{B}(\mathcal{X})$. This is in particular the case for quantile and rank morphisms obtained by Monge's optimal transportation of section 3.3: if

$$\begin{array}{ccc} P^{\mathbf{S}} & \xrightarrow{\mathcal{I}^{\mathbf{H}}} & P^{\mathbf{X}} \\ P^{\mathbf{S}} & \xleftarrow{\mathcal{I}^{\mathbf{H}^{-1}}} & P^{\mathbf{X}} \end{array}$$

are obtained with some optimal transportation map $\mathbf{H}$ of (27), then

$$\mathcal{R}(B_\tau) = \mathcal{I}^{\mathbf{H^{-1}}}(B_\tau) = 1_{\mathbf{H}(B_\tau)},$$

that is to say, the depth regions in the $\mathcal{X}$ world are the direct image by the optimal transportation map of the depth regions in the $\mathcal{S}$ world.

In the case where the morphisms are not degenerate, one only obtains as object a non binary function, i.e. some object interpretable at best as a random set in a enlarged space, see the discussion in remark 2. Moreover, having "depth areas" which are not deterministic subsets also poses epistemological issues and is a matter of debate, see the discussion in section 4.1.

4 A copula and mass transportation approach to quantile morphism

4.1 A discussion on randomization of statistical functionals and a motivation for a combined copula-Monge approach

The discussion of section 2 and of subsection 3.2 showed the necessity to allow for *random* transformations between probability measures in order to be able to define Quantile and Rank morphisms as reciprocals of each other. However, such a stance, mathematically legitimated on the abstract algebraic grounds of category theory, may be objectionable from an epistemological/statistical point of view. Indeed, quoting Cencov [4] p. 6, "The decision-making procedure $\Pi(\omega, de)$ requires that, after observing the outcome ω, an additional, independent, random choice of the inference e be made, based on the law $\Pi(\omega;.)$. This random answer is then a statistical decision by the rule Π." Consequently, switching from the classical viewpoint of statistics $f(\omega)$ as measurable functions of the observations ω to the Blackwell-Le Cam-Cencov theory of statistical inference based on randomized procedure (i.e. Markov kernels $\Pi(\omega, de)$) may be considered problematic from the scientific viewpoint: two statisticians, having the *same* observation ω, with the *same* non degenerate decision rule $\Pi(\omega;.)$, may obtain two different answers on the inference considered. In other words, if used improperly, randomized statistical inference procedures may fail to abide by one of the main criteria of the scientific method, i.e. reproducibility and objectivity of its conclusions in face of common empirical evidence. (Note, however, that extraneous randomisation appears in disguise in several statistical procedure like smoothing, regularization, which may appear "deterministic" at first glance, see the discussion of section 3.2).

One is confronted with an issue similar when one endorses the subjectivist/ Bayesian interpretation of probability: introducing arbitrary or subjective a-priori randomness is likely to introduce arbitrary and subjective conclusion (in finite samples), unless a Bernstein-Von Mises type theorem can come to our rescue and allows us to recover (asymptotically) the true value of the parameter under investigation, whose ontological (and not merely mathematical) existence is posited. Since our intent is not to stir controversy, nor to get to far in "philosophical" discussions, we let the interested reader refer to Bunge [3] for an examination of those issues. Let us mention that a comparable problem can occur in copula theory: by defining a copula attached to a discrete vector $\mathbf{X}$ by the randomised distributional transform (10), one can twist the dependence structure of the copula representer $\mathbf{U}$ corresponding to $\mathbf{X}$, by choosing a randomiser $\mathbf{V}$ with a distribution different from the independence copula, see [12, 42].

In this respect, in order to minimise the subjectivity/perturbation introduced by extraneous randomization, we advocate that Quantile and Rank morphisms should be based, as far as possible, on deterministic transforms (a credo which might seem paradoxical at first glance with the content of section 3). To this end, we propose a Quantile morphism which combines the copula view of section 2 and the Monge transportation approach of section 3. It builds on a mass transportation approach to depth functions by Chernozhukov et al. [5], which we now present.

4.2 The Monge transportation based depth of Chernozukhov et al. [5]

In Chernozukhov et al. [5], a (Monge) transportation approach to quantiles and depths functions is proposed. Their basic idea is that in a spherical distribution, balls give a natural definition of a region which is central for the distribution and which contains most of its mass. Therefore, their basic device is to transform a multivariate $\mathbf{X} \sim F$ into a $\mathbf{S} := (r, \mathbf{a}) \sim P^{\mathbf{S}}$ with spherical uniform distribution on the unit ball $B_1 := \{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}|| \leq 1\}$ of $\mathbb{R}^d$, and conversely,

$$P^{\mathbf{X}} \xrightarrow{\mathcal{R}_F} P^{\mathbf{S}}$$
$$P^{\mathbf{X}} \xleftarrow{\mathcal{Q}_F} P^{\mathbf{S}},$$

with $\mathcal{R}_F = \mathcal{I}^{\mathbf{R}_F}$ and $\mathcal{Q}_F = \mathcal{I}^{\mathbf{Q}_F}$. More precisely, their scheme is as follows:

- One transforms ("polarizes") $\mathbf{X}$ into $\mathbf{S} := (r, \mathbf{a})$, where r stands for a radius uniformly distributed, $r \sim U_{[0,1]}$, and $\mathbf{a}$ for an angle vector, also uniformly distributed on the unit sphere of $\mathbb{R}^d$, with r and $\mathbf{a}$ mutually independent. In transportation theory terms, one transforms $P^{\mathbf{S}}$ into $P^{\mathbf{X}}$ and conversely, via a pair of (deterministic) Monge transformation maps $\mathbf{Q}_F$ and $\mathbf{R}_F$ s.t.

$$\mathbf{Q}_F \# P^{\mathbf{S}} = P^{\mathbf{X}}, \quad \mathbf{R}_F \# P^{\mathbf{X}} = P^{\mathbf{S}}.$$

- One then computes the depth region of content τ on this spherical uniform distribution $\mathbf{S}$: it is simply the ball B_τ of radius τ, since $P^{\mathbf{S}}(B_\tau) = \tau$.

- One back transforms the ball B_τ of radius τ to the original space $\mathbb{R}^d$ where $\mathbf{X}$ lives, via the transformation $\mathbf{Q}_F$:

$$A(\tau, F) := \mathbf{Q}_F(B_\tau).$$

  One then obtains a depth region $A(\tau, F)$ of the kind (7), whose $P^{\mathbf{X}}$ probabilistic content is τ:

$$P^{\mathbf{X}}(A(\tau, F)) = P(\mathbf{Q}_F(\mathbf{S}) \in \mathbf{Q}_F(B_\tau)) = P^{\mathbf{S}}(B_\tau) = \tau$$

- Depth measures for $\mathbf{X}$ can be transfered from depth measures for $\mathbf{S}$ (e.g. Tukey's depth): $D(\mathbf{x}, F) = D^{\text{Tukey}}(\mathbf{R}_F(\mathbf{x}), P^{\mathbf{S}})$

The empirical versions are defined similarly: for samples $\mathbf{X}_1, \mathbf{X}_2, \ldots$ (respectively $\mathbf{S}_1, \mathbf{S}_2, \ldots$) of copies distributed according to F (resp. $P^{\mathbf{S}}$), let F_n (resp. $F_n^{\mathbf{S}}$) the corresponding ecdf. Several variants are proposed, depending on whether one use smoothed versions $\hat{F}_n, \hat{F}_n^{\mathbf{S}}$ of the ecdfs $F_n, F_n^{\mathbf{S}}$, and whether one uses the sample $\mathbf{S}_1, \mathbf{S}_2, \ldots$ or a fixed reference $P^{\mathbf{S}}$ distribution.

4.3 A combined approach

We propose to elaborate on the previous approach of section 4.2, by composing it with a preliminary step of reduction to the copula representation of section 2.2.2. More precisely, the proposed scheme is 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 satisfy Sklar's identity (13).

2. Transport $P^{\mathbf{U}}$ into a spherical distribution $P^{\mathbf{S}}$, via transportation maps $\mathbf{Q}_C, \mathbf{R}_C$;

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

3. Compute the depths regions $A(\tau, P^{\mathbf{S}})$ of level τ, i.e. balls B_τ of radius τ: $A(\tau, P^{\mathbf{S}}) := B_\tau$ and $P^{\mathbf{S}}(A(\tau, P^{\mathbf{S}})) = P^{\mathbf{S}}(B_\tau) = \tau$;

4. Use the transportation maps $\mathbf{Q}_C, \mathbf{R}_C$ to turn these balls into depth regions $A(\tau, P^{\mathbf{U}})$ of level τ at the copula level, i.e. for $\mathbf{U}$;

$$A(\tau, C) := \mathbf{Q}_C(B_\tau), \quad P^{\mathbf{U}}(A(\tau, C)) = \tau$$

5. Use the multivariate marginal quantile transform $\mathbf{G}^{-1}$ to obtain depth regions

$$A(\tau, P^{\mathbf{X}}) = \mathbf{G}^{-1}(A(\tau, P^{\mathbf{U}}))$$

   for the original variable $\mathbf{X}$: $P^{\mathbf{X}}(A(\tau, P^{\mathbf{X}})) = \tau$.

In other words, one defines a pair of Quantile and Rank Morphism $(\mathcal{Q}_{\mathbf{X}}, \mathcal{R}_{\mathbf{X}})$ as,

$$\mathcal{Q}_{\mathbf{X}} := \mathcal{I}^{\mathbf{Q}_C} \circ \mathcal{I}^{\mathbf{G}^{-1}}, \quad \mathcal{R}_{\mathbf{X}} := \mathcal{D}_{\mathbf{X}} \circ \mathcal{I}^{\mathbf{R}_C}, \tag{28}$$

where $\mathcal{D}_{\mathbf{X}}$ is the distributional transform Markov morphism of (24), according to the following diagram,

$$\begin{array}{ll} P^{\mathbf{S}} & \xrightarrow{\mathcal{I}^{\mathbf{Q}_C}} P^{\mathbf{U}} \xrightarrow{\mathcal{I}^{\mathbf{G}^{-1}}} P^X \\ P^{\mathbf{S}} & \xleftarrow{\mathcal{I}^{\mathbf{R}_C}} P^{\mathbf{U}} \xleftarrow{\mathcal{D}_{\mathbf{X}}} P^X. \end{array}$$

The empirical version is defined similarly:

1. conditionally on the sample $\mathbf{X}_1(\omega), \mathbf{X}_2(\omega), \ldots$, set (one bootstrap replication) $\mathbf{X}_n^* \sim F_n$, and define as in (15), $\mathbf{U}_n := \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) \sim C_n$, where C_n is the empirical copula function;
2. transport $P^{\mathbf{U}_n}$ into a spherical distribution $P^{\mathbf{S}}$, via transportation maps $\mathbf{Q}_{C_n}, \mathbf{R}_{C_n}$;
$$\mathbf{Q}_{C_n} \# P^{\mathbf{S}} = P^{\mathbf{U}_n}, \quad \mathbf{R}_{C_n} \# P^{\mathbf{U}_n} = P^{\mathbf{S}}$$
3. the rest of the procedure is the same: one obtain empirical depth area of content τ, as
$$A(\tau, C_n) := \mathbf{Q}_{C_n}(B_\tau), \quad P^{\mathbf{U}_n}(A(\tau, C_n)) = \tau$$
and
$$A(\tau, P^{\mathbf{X}_n^*}) = \mathbf{G}_n^{-1}(A(\tau, C_n)).$$

Symbolically, in terms of morphisms,

$$\begin{array}{ll} P^{\mathbf{X}_n^*} & \xrightarrow{\mathcal{D}_{\mathbf{X}_n^*}} P^{\mathbf{U}_n} \xrightarrow{\mathcal{I}^{\mathbf{R}_{C_n}}} P^{\mathbf{S}} \\ P^{\mathbf{X}_n^*} & \xleftarrow{\mathcal{I}^{\mathbf{G}_n^{-1}}} P^{\mathbf{U}_n} \xleftarrow{\mathcal{I}^{\mathbf{Q}_{C_n}}} P^{\mathbf{S}}, \end{array}$$

where $\mathcal{D}_{\mathbf{X}_n^*}$ is the distributional transform Markov morphism of (24) for $\mathbf{X}_n^*$ into $\mathbf{U}_n$, i.e. the empirical Quantile and Rank Morphisms are

$$\mathcal{Q}_n := \mathcal{I}^{\mathbf{Q}_{C_n}} \mathcal{I}^{\mathbf{G}_n^{-1}}, \quad \mathcal{R}_n := \mathcal{D}_{\mathbf{X}_n^*} \mathcal{I}^{\mathbf{R}_{C_n}} \tag{29}$$

4.4 Discussion

The advantage of such combined copula transportation approach to quantile and depth areas is fourfold:

1. first, as with many copula approaches, standardizing the marginals to uniform distributions on $[0, 1]$ allows to separate the randomness in $P^{\mathbf{X}}$ pertaining to the marginals $\mathbf{G}$ alone from the randomness pertaining to the “dependence” C only. Hence, one reduces the computation of depth

regions only to the dependence structure $P^{\mathbf{U}}$ of $P^{\mathbf{X}}$. Moreover, as copulas are invariant w.r.t. monotone increasing transformations of the marginals, the corresponding depth regions on $\mathbf{U}$ computed from the balls of $P^{\mathbf{S}}$ are invariant w.r.t. monotone increasing transformations of the marginals: combined with the multivariate quantile transforms $\mathbf{G}^{-1}$, one obtains a Markov morphism which automatically satisfy the axiom [O2] of section 3.2. In addition, the corresponding depth measures at the copula level $P^{\mathbf{U}}$ will obey an axiom of monotone invariance which is much more stronger and natural that the axiom (A1) of affine invariance in [5, 47, 48].

2. second, the adjunction of a (continuous, non singular) randomizer $\mathbf{V}$ in the copula transformation step smoothes the empirical copula function C_n (it is even at least Lipshitz). Hence, the empirical copula measure $P^{\mathbf{U}_n}$ is absolutely continuous w.r.t to the d-variate Lebesgue measure λ^d. In addition, since $\mathbf{U}_n, \mathbf{U} \in [0,1]^d$, moments of all order exist. Therefore, the assumptions of Brenier and McCann (theorem 2.1 in [5]) on the existence and unicity of a transportation map are automatically satisfied. The consequence in terms of Markov morphisms is clear: all morphisms in (28) and (29) are of the degenerate type (23), except for the copula morphism $\mathcal{I}^{\mathbf{G}(.,\mathbf{V})}$ and its empirical companion (One obviously chooses a randomizer $\mathbf{V}$ made of independent uniform components, in order not to modify the dependence structure of $\mathbf{X}, \mathbf{X}_n^*$, when viewed in the copula world, through $\mathbf{U}, \mathbf{U}_n^*$). Hence, the proposed combined Quantile and Rank morphism proposed are in agreement with the credo of section 4.1. Moreover, one comply with the assumptions of the powerful theorems 3.2 and 3.4 by Cuesta-Albertos et al. [6], which will prove expedient for the asymptotic analysis of section 5.

3. third, one obtains a smoothing device of the empirical (copula) measure which does not rely on ad-hoc bandwidth parameters as in the classical kernel smoothing approach in [5]: this is relevant from the finite sample point of view, since it is well known that the classical kernel smoothed empirical measure is biased. Hence the resulting transportation maps and depths of [5] are likely to be also biased in finite sample, and one has to optimize the bandwidth in practice.

4. fourth, one obtains a unified approach for both a discrete or a continuous $\mathbf{X}$, and one can therefore extend depth areas to multivariate discrete distributions and get rid of the continuity assumptions in [5].

5 Asymptotic results

The proofs of the consistency in probability of transportation maps and depth measures in [5] (their theorem 3.1) are analytical and are based on results on the local uniform convergence of subdifferentials, via duality analysis. As in Faugeras [11, 12], we favor the use of the method of a single probability space (a.s. constructions, see Skorohod [54], Varadarajan [58], Ranga-Rao [33]) which allows to use results by Cuesta-Albertos et al. (theorem 3.4 in [6]), Ranga Rao ([33]), and Faugeras ([11, 12]).

5.1 Framework

The setting is as follows:

- Framework:

  Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ an infinite dimensional sample defined, w.l.o.g. on the canonical countably infinite product probability space

  $$(\Omega, \mathcal{A}, P) := (\mathbb{R}^d \times \mathbb{R}^d \times \ldots, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \ldots, P).$$

  In other words $\omega = (\omega_1, \omega_2, \ldots)$ and $\mathbf{X}_i(\omega) = \omega_i$, the coordinate projections. Let P_n the empirical measure based on the $n-$sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$,

  $$P_n(.) := P_n^{\omega}(.) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\mathbf{X}_i(\omega)}(.),$$

  and F_n, $\mathbf{G}_n$ its c.d.f. and corresponding vector of marginal e.c.d.f. Such $P_n(.) = P_n^{\omega}(.)$ is a random measure, i.e. it can be construed as a Markov kernel from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, (and we suppress the dependence on ω, as is customary).

- Ergodicity hypothesis:

  Assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ is an ergodic sample of $P^{\mathbf{X}}$, in the sense that, for each real-valued function g on $\mathbb{R}^d$, s.t. $P^{\mathbf{X}}(|g|) = \int |g| dP^{\mathbf{X}} < \infty$,

  $$P_n(g) := \int g(\mathbf{x}) P_n(d\mathbf{x}) \to P^{\mathbf{X}}(g), \quad P- \text{ almost everywhere} \tag{30}$$

  Remark 5. *The above definition is a specialization of Ranga-Rao's [33] definition of ergodicity to a non-random invariant measure, similar to definition (E) of [5] and the framework of [12].*

- Assumption on $P^{\mathbf{X}}$:

  We will also make one of the following assumption:

  - (C): $P^{\mathbf{X}} << \lambda^d$,
  - (D): $P^{\mathbf{X}}$ is discrete.

5.2 Main theorem

Theorem 5.1. *Assume (30) and either (C) or (D). Then with P-probability one, one can construct on some probability space $(\Omega^*, \mathcal{A}^*, P^*)$ a sequence $\mathbf{S}_n, \mathbf{S}$ of random vector distributed as $P^{\mathbf{S}}$, and a sequence of copula representers $\mathbf{U}_n, \mathbf{U}$*

distributed as C_n, C, a sequence of random vectors $\mathbf{X}_n^*, \mathbf{X}^*$ distributed as F_n, F s.t., with $P-$probability one,

$$(\mathbf{U}_n, \mathbf{S}_n) \stackrel{d}{\to} (\mathbf{U}, \mathbf{S}),$$

and also, with $P-$probability one,

$$(\mathbf{X}_n^*, \mathbf{U}_n, \mathbf{S}_n) \to (\mathbf{X}^*, \mathbf{U}, \mathbf{S}) \quad P^* - a.s.$$

Proof.

- Step one: ergodicity imply weak convergence of empirical measures with probability one.

  By Varadajan-Ranga Rao's extension of the Glivenko-Cantelli's theorem (See Ranga-Rao [33] Theorem 6.1 and [58]), (30) implies

  $$P[P_n \stackrel{d}{\to} P^{\mathbf{X}}] = 1 \tag{31}$$

  In other words, there exists $\Omega_0 \subset \Omega$, with $P(\Omega_0) = 1$, s.t. for all $\omega \in \Omega_0$, $P_n \stackrel{d}{\to} P^{\mathbf{X}}$. Pick some fixed $\omega \in \Omega_0$.

- Step two: a.s. convergence of copula representers on a suitable probability space.

  On some (uninteresting) extra probability space $(\Omega^*, \mathcal{A}^*, P^*)$, define, conditionally on ω, one (bootstrap) representing variable $\mathbf{X}_n^* : \Omega^* \to \mathbb{R}^d$ of P_n, for each $n \in \mathbb{N}^*$, i.e. $\mathbf{X}_n^* \sim P_n$. Define similarly $\mathbf{X}^* : \Omega^* \to \mathbb{R}^d$ in such a conditional manner s.t. its law be $P^{*\mathbf{X}^*} = P^{\mathbf{X}}$. (We have dropped the dependency of these random elements on the chosen $\omega \in \Omega_0$). Note that such conditional measures are guaranteed to exists and to be genuine probability measures, since the underlying spaces are Polish. In the remaining, we will also suppress "with P probability one" in our statements, corresponding to the fact that $\omega \in \Omega_0$ with $P(\Omega_0) = 1$ according to (31). By Skorohod's theorem, $(\Omega^*, \mathcal{A}^*, P*)$, $\mathbf{X}_n^*, \mathbf{X}^*$ can be chosen so that

  $$\mathbf{X}_n^* \stackrel{P^* a.s.}{\longrightarrow} \mathbf{X}^*$$

  Set $\mathbf{V}$, defined also on $(\Omega^*, \mathcal{A}^*)$, a vector with uniform marginals, independent of $(\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}^*)$ (Enlarge the probability space by product if necessary). Set

  $$\mathbf{U}_n := \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) \sim C_n$$

  so that $\mathbf{X}_n^* = \mathbf{G}_n^{-1}(\mathbf{U}_n)$. Similarly, set $\mathbf{U} := \mathbf{G}(\mathbf{X}^*, \mathbf{V})$.

  If assumption (D) is true (i.e. when $P^{\mathbf{X}}$ is discrete), then $\mathbf{U}_n \stackrel{P^*-a.s.}{\to} \mathbf{U}$, thanks to the almost sure convergence theorem of the empirical copula representer for an ergodic sample (see theorem 3.1 by Faugeras [12]).

  If assumption (C) is true, then F is continuous and a.s. consistency of copula representer is an easy consequence of Skorohod's theorem, as shown in the the following lemma, whose proof is relegated in the appendix:

Lemma 5.2. *If F is continuous, then* $\mathbf{U}_n \overset{P^* a.s.}{\rightarrow} \mathbf{U}$.

- Step three: a.s. convergence of copula-Monge transportation representers

  As proposed in section 4.3, transport $P^{\mathbf{U}_n}$ towards the reference spherical distribution $P^{\mathbf{S}}$ and conversely by solving the Monge-Kantorovich problem with quadratic cost (Wasserstein distance),

$$
\begin{aligned}
P^{\mathbf{U}_n} &\quad \overset{\mathcal{Q}_{C_n}}{\longleftarrow} P^{\mathbf{S}}, \\
P^{\mathbf{U}_n} &\quad \overset{\mathcal{R}_{C_n}}{\longrightarrow} P^{\mathbf{S}},
\end{aligned}
$$

  where the Markov morphism, as solution of Monge's problem with quadratic cost (see section 3.3) are of degenerate type (see definition 3.3 or (23)), i.e. are induced by the transportation maps $\mathbf{Q}_{C_n}$ and $\mathbf{R}_{C_n}$,

$$\mathcal{Q}_{C_n} = \mathcal{I}^{\mathbf{Q}_{C_n}}, \quad \mathcal{R}_{C_n} = \mathcal{I}^{\mathbf{R}_{C_n}}.$$

  Equivalently, in the push-forward notation,

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

  Indeed, since both distributions $P^{\mathbf{U}_n}$ and $P^{\mathbf{S}}$ have compact support and are absolutely continuous, the assumptions of Brenier and Mc Cann's theorem are satisfied and the transportation maps $\mathbf{Q}_{C_n}$ and $\mathbf{R}_{C_n}$ exist and are unique. Similarly, assumption (C) or (D) yields that the transportation maps

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

  also exists and are unique.

  Hence, one can realize these distributions on $(\Omega^*, \mathcal{A}^*)$ by defining the random vectors $\mathbf{S}_n$ and $\mathbf{S}$, distributed as $P^{\mathbf{S}}$, by

$$\mathbf{S}_n := \mathbf{R}_{C_n}(\mathbf{U}_n), \quad \mathbf{S} := \mathbf{R}_C(\mathbf{U})$$

  which also satisfy

$$\mathbf{Q}_{C_n}(\mathbf{S}_n) = \mathbf{U}_n, \quad \mathbf{Q}_C(\mathbf{S}) = \mathbf{U}.$$

  Now, the assumptions of theorems 3.2 and 3.4 in Cuesta et al. [6] (see also Theorem 3.2 in Tuero [57]) are satisfied. Hence, $\mathbf{U}_n \overset{P^* a.s.}{\rightarrow} \mathbf{U}$ yields

$$(\mathbf{U}_n, \mathbf{S}_n) \overset{d}{\rightarrow} (\mathbf{U}, \mathbf{S})$$

  and

$$\mathbf{S}_n := \mathbf{R}_{C_n}(\mathbf{U}_n) \overset{P^* a.s.}{\rightarrow} \mathbf{R}_C(\mathbf{U}) := \mathbf{S} \tag{32}$$

  Thanks to (31), the latter results are true with P-probability one.

  □

5.3 Convergence of depth areas in average symmetric difference distance

The (population and empirical) depth area of mass τ are defined from the centered ball B_τ of radius τ via the combined Markov morphism $\mathcal{Q}_{\mathbf{X}}$ of (28) and $\mathcal{Q}_n$ of (29) as

$$\begin{aligned} A &:= \mathcal{Q}_{\mathbf{X}} B_\tau = \mathcal{D}_{\mathbf{X}} \mathcal{I}^{\mathbf{R}_C} B_\tau \\ A_n &:= \mathcal{Q}_n B_\tau = \mathcal{D}_{\mathbf{X}_n^*} \mathcal{I}^{\mathbf{R}_{C_n}} B_\tau \end{aligned}$$

which reduces, due to the degeneracy of the Markov morphisms, to the sets

$$\begin{aligned} A &:= \mathbf{G}^{-1} \circ \mathbf{Q}_C(B_\tau) \\ A_n &:= \mathbf{G}_n^{-1} \circ \mathbf{Q}_{C_n}(B_\tau) \end{aligned}$$

A way to measure the distance between these sets is through their average symmetric difference

$$P^{\mathbf{X}}(A_n \Delta A) = P^{\mathbf{X}} |1_{A_n} - 1_A|$$

(which generalizes to the L_1 distance w.r.t. $P^{\mathbf{X}}$ when A_n, A are functions).

Corollary 5.3. *With P-probability one, the $L_1(P^{\mathbf{X}})$ or symmetric difference distance between the $P^{\mathbf{X}}$ depth area and its empirical counterpart converges towards zero, as $n \to \infty$,*

$$P^{\mathbf{X}}(A_n \Delta A) \to 0.$$

Proof. By definition of the random variables of the previous subsection, $\mathbf{X}^* \in A \Leftrightarrow \mathbf{X}^* \in \mathbf{G}^{-1} \circ \mathbf{Q}_C(B_\tau) \Leftrightarrow \mathbf{S} \in B_\tau$ and $\mathbf{X}^* \in A_n \Leftrightarrow \mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}^*, \mathbf{V}) \in B_\tau$, P^* a.s.

Hence,

$$\begin{aligned} P^{\mathbf{X}}(A_n \Delta A) &= P^* |1_{\mathbf{S} \in B_\tau} - 1_{\mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}^*, \mathbf{V}) \in B_\tau}| \\ &= P^*(\mathbf{S} \in B_\tau, \mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}^*, \mathbf{V}) \notin B_\tau) \\ &+ P^*(\mathbf{S} \notin B_\tau, \mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}^*, \mathbf{V}) \in B_\tau) \end{aligned}$$

By the continuous mapping theorem, $\mathbf{R}_{C_n} \circ \mathbf{G}_n(\mathbf{X}^*, \mathbf{V}) \xrightarrow{P^* a.s.} \mathbf{S}$, hence the above two probabilities go to zero as $n \to \infty$, since B_τ is a $P^{\mathbf{S}}$ continuity set. □

Remark 6. *The $P^{\mathbf{X}}$-L_1 or $P^{\mathbf{X}}$-averaged symmetric difference distance is well suited to the problem at hand, whereas the Hausdorff distance of [15], being intrinsic, mandates special restrictive conditions to the ranges of the c.d.fs. and also to avoid infinities. For completeness, let us briefly sketch an argument for convergence in the Hausdorff metric: by corollary A.1 in [5], it suffices to show uniform convergence of*

$$\mathbf{G}_n^{-1} \circ \mathbf{Q}_{C_n} \to \mathbf{G}^{-1} \circ \mathbf{Q}_C \tag{33}$$

on some suitable compact subsets of the unit ball B_1. Uniform convergence on compacta of optimal mappings $\mathbf{Q}_{C_n} \to \mathbf{Q}_C$ is provided by checking conditions of their theorems A.1 and A.2: in the notation of [5], $\mathcal{U} = B_1$, $\mathcal{Y} = [0,1]^d$ are compact and convex; $P^{\mathbf{S}}, P^{\mathbf{U}_n}$ are absolutely continuous w.r.t. $d-$variate Lebesgue measure, under the choice of of a coordinate independent randomizer $\mathbf{V}$ (see section 2.2.2) and under assumption (C) or (D) of our paper, $P^{\mathbf{U}}$ is also absolutely continuous w.r.t. $d-$variate Lebesgue measure; condition (W) in theorem A.2 of [5] holds by theorem 5.1; condition (C) in theorem A.2 of [5] holds for optimal gradient mappings $(\mathbf{Q}_C, \mathbf{R}_C)$ and $(\mathbf{Q}_{C_n}, \mathbf{R}_{C_n})$ on the sets $\mathcal{U}_0 = int(suppP^{\mathbf{S}})$ and $\mathcal{Y}_0 = int(suppP^{\mathbf{U}}) = (0,1)^d$. Hence, one obtains uniform convergence of $\mathbf{Q}_{C_n} \to \mathbf{Q}_C$ on compact subsets K of $\mathcal{U}_0$, and that $d_H(\mathbf{Q}_{C_n}(B_\tau), \mathbf{Q}_C(B_\tau)) \to 0$, with $P-$ probability one.

Next, one needs to prove uniform or continuous convergence of

$$\mathbf{G}_n^{-1} \to \mathbf{G}^{-1}. \tag{34}$$

The most simple case is when $\mathbf{G}^{-1}$ is strictly increasing on its domain. Then, (34) holds pointwise everywhere and uniform convergence holds on $\{|\mathbf{G^{-1}}| \le \mathbf{k}\}$ or any $\mathbf{k} < \infty$ and (33) holds uniformly on $K \cap \{\mathbf{R}_C(|\mathbf{G^{-1}}| \le \mathbf{k})\} =: K_{\mathbf{k}}$, where $K \subset \mathcal{U}_0$. In the general case, under (C) or (D), the discontinuity set D of $\mathbf{G}^{-1}$ is at most "grid" parallel to the coordinate axis of dimension lower than $d-1$, and we can replace $K_{\mathbf{k}}$ above by $K'_{\mathbf{k}} := K \cap \{\mathbf{R}_C(|\mathbf{G^{-1}}| \le \mathbf{k} \cap K_0)\}$ where $K_0 \subset D^c$ is any compact subset. One then obtains (33) on $K_{\mathbf{k}}$ or $K'_{\mathbf{k}}$ and Hausdorff convergence of the depth areas restricted on those sets, with $P-$probability one.

6 Appendix

6.1 Proof of lemma 5.2

Proof. Under Assumption (C), (31) imply the usual Polya-Glivenko-Cantelli theorem, $||F_n - F||_\infty \to 0$, and $||\mathbf{G}_n - \mathbf{G}||_\infty \to 0$, by Ranga-Rao's theorem 4.1 (see also his theorem 3.4) applied to the coordinate projections $l_i(\mathbf{x}) = x_i$.

One has the decomposition,

$$\begin{aligned} \mathbf{U}_n - \mathbf{U} &= \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) - \mathbf{G}(\mathbf{X}_n^*, \mathbf{V}) + \mathbf{G}(\mathbf{X}_n^*, \mathbf{V}) - \mathbf{G}(\mathbf{X}) \\ &= \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) - \mathbf{G}(\mathbf{X}_n^*) + \mathbf{G}(\mathbf{X}_n^*) - \mathbf{G}(\mathbf{X}) \end{aligned}$$

since $\mathbf{G}$ is continuous. But

$$\mathbf{G}_n(\mathbf{X}_n^*-) \le \mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) \le \mathbf{G}_n(\mathbf{X}_n^*)$$

where $\mathbf{G}(\mathbf{x}-)$ denotes the left-hand limit of $\mathbf{G}$. Therefore,

$$|\mathbf{G}_n(\mathbf{X}_n^*, \mathbf{V}) - \mathbf{G}(\mathbf{X}_n^*)| \le ||\mathbf{G}_n - \mathbf{G}||_\infty + \mathbf{1/n}$$

By Skorohod's theorem, $F_n \stackrel{d}{\to} F$ imply $\mathbf{X}_n^* \stackrel{a.s.}{\to} \mathbf{X}$ for some copies on some probability space. Hence, by the continuous mapping theorem $\mathbf{G}(\mathbf{X}_n^*) \to \mathbf{G}(\mathbf{X})$

and

$$||\mathbf{U}_n - \mathbf{U}|| \to 0,$$

as $n \to \infty$. $\square$

6.2 A reminder on convergence of o.t.p.

For the convenience of the reader, we recall below the main result of Cuesta-Albertos et al. [6]: let $\mathcal{P}_2$ the set of probability measures on $\mathbb{R}^d$ with finite second moment: $\mathcal{P}_2 = \{P : \int ||\mathbf{x}||^2 dP < \infty\}$.

Theorem 6.1 (Theorem 3.2 by [6]). *Let $(P_n)_{n\in\mathbb{N}}$, $(Q_n)_{n\in\mathbb{N}}$, P, Q be probability measures in $\mathcal{P}_2$ such that $P << \lambda^d$ and such that $P_n \xrightarrow{w} P$ and $Q_n \xrightarrow{w} Q$. Let us assume that $\mathbf{T}_n$ (resp. $\mathbf{T}$) are o.t.p.'s between P_n and Q_n (resp. P and Q), $n \in \mathbb{N}$. Then,*

$$(\mathbf{X}_n, \mathbf{T}_n(\mathbf{X}_n)) \xrightarrow{d} (\mathbf{X}, \mathbf{T}(\mathbf{X})).$$

Theorem 6.2 (Theorem 3.4 by [6]). *With the same assumptions as in theorem 6.1, if, in addition, $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is a sequence of r.v.'s which converges a.s. and $\mathbf{X}_n \sim P_n$, we have that, almost surely,*

$$\mathbf{T}_n(\mathbf{X}_n) \to \mathbf{T}(\mathbf{X}).$$

References

[1] Jean-Pierre Aubin and Hélène Frankowska. *Set Valued Analysis.* Birkhauser, 1990.

[2] T.S. Blyth. *Lattices and Ordered Algebraic Structures.* Springer, 2005.

[3] Mario Bunge. Two faces and three masks of probability. In Evandro Agazzi, editor, *Probability in the Sciences, Synthese Library*, volume 201, pages 27–50, 1988.

[4] N. N. Cencov. *Statistical Decision Rules and Optimal Inference*, volume 53. Translations of Mathematical monographs, AMS, 1982.

[5] Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-kantorovich depth, quantiles, ranks and signs. *AOS*, 2015. Forthcoming, cemmap working paper CWP04/15.

[6] 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.

[7] J. A. Cuesta-Albertos, L. Rüschendorf, and A. Tuero-Diaz. Optimal coupling of multivariate distributions and stochastic processes. *Journal of Multivariate Analysis*, 46:335–361, 1993.

[8] B. A. Davey and H. A. Priestley. *Introduction to Lattice and Orders*. Cambridge university press, 2nd edition, 2002.

[9] Paul Deheuvels. La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique*, 65(f.6):274–292, 1979.

[10] Paul Deheuvels. A multivariate Bahadur-Kiefer representation for the empirical copula process. *Journal of Mathematical Sciences*, 163:382–398, November 2009.

[11] Olivier P. Faugeras. Sklar's theorem derived using probabilistic continuation and two consistency results. *Journal of Multivariate Analysis*, 122:271 – 277, 2013.

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

[13] Jean-David Fermanian, Dragan Radulovic, and Marten Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860, October 2004.

[14] Warren Gilchrist. *Statistical Modelling with Quantile Functions*. Chapman and Hall/CRC, 2000.

[15] Marc Hallin, Davy Paindaveine, and Miroslav Siman. Multivariate quantiles and multiple-output regression quantiles: From l1 optimization to halfspace depth. *The Annals of Statistics*, 38(2):635–669, 04 2010.

[16] Olav Kallenberg. *Foundations of Modern Probability*. Springer, 1997.

[17] S.M. Lane and G. Birkhoff. *Algebra*. AMS Chelsea Publishing Series. Chelsea Publishing Company, 1999.

[18] William F. Lawvere and Robert Rosebrugh. *Sets for Mathematics*. Cambridge University Press, 2003.

[19] William F. Lawvere and Stephen H. Schanuel. *Conceptual Mathematics*. Cambridge University Press, 1997.

[20] William A. Massey. Stochastic orderings for markov processes on partially ordered spaces. *Mathematics of operation research*, 12(2):350–367, 1987.

[21] David S. Moore and M. C. Spruill. Unified large-sample theory of general chi-squared statistics for tests of fit. *The Annals of Statistics*, 3(3):599–616, May 1975.

[22] Alfred Müller. Stochastic orders generated by integrals: a unified study. *Adv. Apppl. Prob.*, 29:414–428, 1997.

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

[24] K. R. Parthasarathy. *Probability Measures On Metric Spaces.* Academic press, 1967.

[25] Emanuel Parzen. Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74(365):105–121, 1979.

[26] Emmanuel Parzen. Quantile functions, convergence in quantile, and extreme value distribution theory. Technical Report B-3, Statistical Institute, Texas A & M University, College Station, Texas, 1980.

[27] M. B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis.* Academic Press, 1989.

[28] S. T. Rachev. The monge-kantorovich mass transference problem and its stochastic applications. *Theory of Probability and its Applications*, 29(4):647–676, 1985.

[29] S.T. Rachev and L. Rüschendorf. A counterexample to a.s. constructions. *Statistics and Probability Letters*, 9(4):307–309, April 1990.

[30] Svetlozar T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York, 1991.

[31] Svetlozar T. Rachev, Lev B. Klebanov, Stoyan V. Stoyanov, and Frank J. Fabozzi. *The Methods of Distances in the Theory of Probability and Statistics.* Springer, 2013.

[32] Svetlozar T Rachev and Ludger Rüschendorf. *Mass Transportation Problems: Volume I: Theory*, volume 1. Springer Science & Business Media, 1998.

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

[34] Yann Renier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications in Pure and Applied Mathematics*, 44:375–417, 1991.

[35] T.W. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, 1998.

[36] Murray Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

[37] Ludger Rüschendorf. Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics*, 4(5):912–923, September 1976.

[38] Ludger Rüschendorf. Stochastically ordered distributions and monotonicity of the oc-function of sequential probability ratio tests. *Math. Operationforsch. Statist., Series Statistics*, 12(3):327–338, 1981.

[39] Ludger Rüschendorf. The wasserstein distance and approximation theorems. *Z. Wahrsscheinlichkeitstheorie verw. Gebiete*, 70:117–129, 1985.

[40] Ludger Rüschendorf. Fréchet-bounds and their applications. In G. Dall'Aglio, S. Kotz, and G. Salinetti, editors, *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, pages 151–187. Springer Netherlands, Dordrecht, 1991.

[41] Ludger Rüschendorf. Monge-kantorovich transportation problem and optimal couplings. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 109(3):113–137, 2007.

[42] Ludger Rüschendorf. On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

[43] Ludger Rüschendorf. *Mathematical Risk Analysis - Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, 2013.

[44] Ludger Rüschendorf and Svetlovar T. Rachev. A characterization of random variables with minimum l2-distance. *Journal of Multivariate Analysis*, 32(1):48–54, 1990.

[45] Bernd S. W. Schröder. *Ordered Sets: an Introduction*. Springer, 1th edition, 2002.

[46] Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002.

[47] Robert Serfling and Y. Zuo. General notion of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000.

[48] Robert Serfling and Y. Zuo. Structural properties of and convergence results for contours of sample statistical depth functions. *The Annals of Statistics*, 28(2):483–499, 2000.

[49] Robert Serfling and Y. Zuo. Discussion. *The Annals of Statistics*, 38(2):676–684, 2010.

[50] A. N. Shiryaev and V. G. Spokoiny. *Statistical Experiments and decisions: asymptotic theory*, volume 8 of *Advanced series on statistical science and applied probability*. World Scientific, 2000.

[51] Galen R. Shorack. *Probability for Statisticians*. Springer, 2000.

[52] Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

[53] Abe Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460, 1973.

[54] A. V. Skorokhod. Limit theorems for stochastic processes. *Theory of Probability and its Applications*, 1:261–290, 1956.

[55] R. Szekli. *Stochastic Ordering and Dependence in Applied Probability.* Lecture Notes in Statistics , Springer-Verlag, 1995.

[56] Hermann Thorisson. *Coupling, Stationarity, and Regeneration.* Springer, 2000.

[57] Araceli Tuero. On the stochastic convergence of representations based on wasserstein metrics. *The Annals of Probability*, 21(1):72–85, 1993.

[58] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhya*, 19(1/2):23–26, February 1958.

[59] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, 2003.

[60] Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.

[61] David Williams. *Probability with Martingales.* Cambridge University Press, 1991.

[62] V. M. Zolotarev. *Modern Theory of Summation of Random Variables.* VSP BV, 1997.