Kendall quantile ordering on R2 and associated empirical quantile transform map

Philippe Berthet, Jean-Claude Fort

KENDALL QUANTILE ORDERING ON $\mathbb{R}^2$ AND ASSOCIATED EMPIRICAL QUANTILE TRANSFORM MAP

A PREPRINT

Philippe Berthet[*] **Jean Claude Fort**[†]

November 6, 2023

ABSTRACT

We introduce new geometrical characteristics of any smooth probability distribution function F on $\mathbb{R}^2$, such as bivariate quantiles and bivariate ranks. For this we build a two dimensional $r.v.$ generator $\mathcal{G}_F$ from the curves and curvilinear measures induced by what we call the Kendall ordering of F. The ensuing quantile transform $\tau_{FG} = \mathcal{G}_F \circ \mathcal{G}_F^{-1}$ of F into G is a closed form, continuous transport map having natural statistical applications. The fact that $\mathcal{G}_F$ and τ_{FG} are explicitly derived from F makes their empirical counterparts $\mathcal{G}_{F,n}$ and $\tau_{n,m}$ fastly computable from samples of F and G of very large size n and m. From the geometrical nature of $\mathcal{G}_{F,n}$ we construct non parametric depth parameters, local depth fields, contours and a bivariate Kendall tau. We also illustrate how $\tau_{n,m}$ can be used to quantify tests - goodness of fit, comparison, same copula or radiality. The paper focusses on basic properties of τ_{FG} then numerically illustrates most of the introduced notions.

Keywords Generalized quantile transform · Bivariate ranks · Empirical distribution · Kendall distribution · Explicit coupling transport map · Depth · Kendall tau.

AMS Subject Classification: 62G30 ; 60E10 ; 62H20 ; 60F15

1 Introduction

The main topic to be addressed is a bivariate generalization of the univariate quantile transform. The notion of quantile points and ordered rank vectors we use rely on a probabilistic geometry induced by the distribution function $(d.f.)$.

1.1 Flat geometry of a univariate distribution

Let introduce on the real line the notions to be generalized in dimension two with the notation of the paper. Let F be a $d.f.$ with positive density f on an interval. The quantile function $\mathcal{G}_F = F^{-1}$ is a universal continuous generator in the sense that $\mathcal{G}_F(Z)$ has $d.f.$ F if, and only if, the $r.v.$ Z has uniform $d.f.$ U on $(0,1)$. Likewise the inverse generator function $\mathcal{G}_F^{-1} = F$ is a universal ordering function, in the sense that the random rank $\mathcal{G}_F^{-1}(X)$ has $d.f.$ U and $z = \mathcal{G}_F^{-1}(x)$ is the order of the deterministic quantile x of F. We shall think of the rank z as the accumulated generator probability mass before reaching x in a continuous mass dropping process starting from $-\infty$. This is an intuitive way to economically transport U onto F, in a push forward manner. This mass ordering presides over the so-called rank to quantile transform map $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$, which is also an optimal transport map. For any non negative convex function c with $c(0) = 0$ and any $d.f.$ G with positive density g on an interval,

$$\min_{X' \sim F, Y' \sim G} \mathbb{E}\, c(X' - Y') = \int_0^1 c(\mathcal{G}_F(z) - \mathcal{G}_G(z))dz \tag{1}$$

so that the minimum is achieved by $Y' = \tau_{FG}(X')$. In other words, the random rank of X' and Y' are always the same, these two random quantiles of F and G have been dropped simultaneously. Hence a probability distribution on $\mathbb{R}$ is not

[*]Institut de Mathématiques de Toulouse UMR 5219 ; Université Paul Sabatier, France. *philippe.berthet@math.univ-toulouse.fr*

[†]MAP5 UMR 8145; Université Paris Cité, France. *jean-claude.fort@parisdescartes.fr*

described by the probability of Borel sets but by points x indexed by the flat geometry of uniform ranks $F(x)$ ordered in the generator space $(0,1)$. This is exactly the way we shall procede in dimension two, with curves replacing points.

In this paper we propose to keep $F(X)$ playing a central role as part of a universal generator dropping the probability mass F continuously. For this we separate out the space of the bivariate generator $r.v.$ Z carrying a universal ordering onto the space of the bivariate $r.v.$'s to be generated, ranked, quantile transformed. Our mass dropping approach relies on the intrinsic probabilistic geometry generated by $(F(X), X|F(X))$, made of two deterministic collections of measured curves. This reduces generating $X = \mathcal{G}_F(Z)$ to intersecting two curves.

1.2 Curves geometry of a bivariate distribution

The class $\mathbb{F}$. Consider two $r.v.$'s X and Y having bounded parallel rectangular supports $\overline{\mathcal{R}}_X$ and $\overline{\mathcal{R}}_Y$ that are the closure of $\mathcal{R}_X = \left(x_1^-, x_1^+\right) \times \left(x_2^-, x_2^+\right)$ and $\mathcal{R}_Y = \left(y_1^-, y_1^+\right) \times \left(y_2^-, y_2^+\right)$. Assume that they have $\mathcal{C}^2$ distribution functions and positive $\mathcal{C}^1$ densities on $\mathcal{R}_X$ and $\mathcal{R}_Y$ denoted (F, f) and (G, g) respectively. Write $\mathbb{F}$ the family of such smooth distributions and $\overline{\mathbb{F}}$ its extension allowing parallel unbounded rectangles, including $\mathbb{R}^2$.

The generator $r.v.$ Let $Z = (Z_1, Z_2)$ have uniform $d.f.$ denoted U on $\mathbb{U} = (0,1)^2$, with support $\overline{\mathbb{U}} = [0,1]^2$.

Generator. We say that a continuous, one to one function $\mathcal{G}_F : \mathbb{U} \to \mathcal{R}_X$ is a generator of $F \in \mathbb{F}$ if $\mathcal{G}_F(Z)$ has $d.f.$ F. The generator $\mathcal{G}_F$ we define at Section 2 relies on what we call the Kendall geometry.

Kendall geometry. The Kendall $d.f.$ was introduced in [8] in a bivariate setting as the $d.f.$ of $F(X)$. This notion was further studied in [23] for copulas, and generalized in higher dimension in [2]. We exploit the geometrical aspects of the Kendall $d.f.$ one step further by studying the conditional probability along the level curves of F, that we call the Q-curves, having probability Z_1 below them. We thus define a second family of curves, that we call the R-curves, to fully characterize F as they yield a region below the Q-curves having probability $Z_1 Z_2$. Any Q-curve intersects any R-curve in a unique point, and this is the mathematically appealing way we define $\mathcal{G}_F(Z)$.

Quantile and rank. We call $\mathcal{G}_F(z) = x$ the bivariate quantile of X of order $z = (z_1, z_2) \in \mathbb{U}$ and $\mathcal{G}_F^{-1}(x) = z$ the bivariate rank of $x = (x_1, x_2) \in \mathcal{R}_X$. Hence we identify the generator coordinates with the rank read in Kendall order.

Kendall ordering. The rank coordinate $z = (z_1, z_2) \in \mathbb{U}$ is endowed with the strict order

$$z < z^+ \text{ if either } z_1 < z_1^+ \text{ or } z_1 = z_1^+ \text{ and } z_2 < z_2^+. \tag{2}$$

Thus $z \leq z'$ if $z < z'$ or $z = z'$. The Kendall ordering is characterized as follows: if $x = \mathcal{G}_F(z)$ then

$$\mathbb{P}(F(X) \leq F(x)) = z_1, \quad \mathbb{P}(X_2 \geq x_2 \mid F(X) = F(x)) = z_2,$$

which, in the quadrant oriented geometry driving $\mathcal{G}_F$, is equivalent to

$$\mathbb{P}(F(X) \leq F(x)) = z_1, \quad \mathbb{P}(X_1 \leq x_1 \mid F(X) = F(x)) = z_2.$$

In other words, for all $z \in \mathbb{U}$ and $F \in \mathbb{F}$ we have

$$z_1 = \mathbb{P}(\mathcal{G}_F^{-1}(X) \leq z), \; z_2 = \mathbb{P}(X_2 \geq \langle \mathcal{G}_F(z), e_2 \rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\} \times (0,1)) \tag{3}$$

and, equivalently, $z_2 = \mathbb{P}(X_1 \leq \langle \mathcal{G}_F(z), e_1 \rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\} \times (0,1))$ where (e_1, e_2) is the usual orthonornal basis of $\mathbb{R}^2$. As a matter of fact, $\mathcal{G}_F^{-1}$ induces the following geometrical and stochastic orders on $\mathbb{F}$.

Quantile ordering. We say that $x \leq x^+$ in the Kendall geometry of F if $\mathcal{G}_F^{-1}(x) \leq \mathcal{G}_F^{-1}(x^+)$ in the sense (2) of the generator coordinates, which means that either $F(x) < F(x^+)$ or

$$F(x) = F(x^+), \; \mathbb{P}(X_2 \geq x_2 \mid F(X) = F(x)) \leq \mathbb{P}(X_2 \geq x_2^+ \mid F(X) = F(x)).$$

For $F, G \in \mathbb{F}$ we can stochastically compare X and Y by ranks, $X \leq Y$ if $\mathcal{G}_F^{-1}(X) \leq \mathcal{G}_G^{-1}(Y)$.

Bivariate quantile transform map. Clearly, the bidimensional quantile transform map

$$\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1} \tag{4}$$

preserves the Kendall quantile ordering on the bivariate rank square $\mathbb{U}$. It is a rank to quantile coupling enjoying partial optimality generalizing (1) – for finite collections of curves, see Section 3. Our main result is the existence and uniqueness of a generator preserving the Kendall quantile ordering.

Theorem 1. *There exists a unique map $\mathcal{G} : F \in \mathbb{F} \to \mathcal{G}(F) = \mathcal{G}_F$ such that for $F \in \mathbb{F}$, $\mathcal{G}_F(Z)$ has distribution F, $\mathcal{G}_F$ satisfies (3) and, for $(F, G) \in \mathbb{F} \times \mathbb{F}$, the quantile transform τ_{FG} of (4) is one to one, continuously differentiable on $\mathcal{R}_X$.*

The map $\mathcal{G}$ is explicitly constructed from the geometry developed at Section 2 – making Section 1.1 precise. It has properties of probabilistic nature (Section 3), statistical nature (Sections 4, 6) and numerical nature (Section 5).

1.3 Empirical quantile transform map

The quantile transform τ_{FG} of (4) can be learned in a non-parametric way by sampling then matching the empirical geometries derived from the respective empirical $d.f.$'s. The key advantage is that the Q-curves and R-curves are efficiently estimated by sorting n sample points in a special way – see Section 5. The empirical $\mathcal{G}_{F,n}$ and $\mathcal{G}_{F,n}^{-1}$ then reveal to be as flexible as in the univariate case, while carrying the joint information beyond marginals. It turns out that $\tau_{n,m} = \mathcal{G}_{G,m} \circ \mathcal{G}_{F,n}^{-1}$ can be computed from samples up to several millions in a reasonable time. Therefore we propose new statistical tools based on the Kendall geometry and bivariate quantile transforms.

From $\mathcal{G}_F$ we define new geometrical characteristics of F such as global and local depth fields and contours fitting the mass localization. The empirical counterparts based on $\mathcal{G}_{F,n}$ are easily computed. Likewise, using $\tau_{n,m}$ we introduce new goodness of fit tests and bivariate samples comparison statistical tools. For instance, testing for same copula or testing for same radiality – same angle distribution. We also introduce a bivariate rank correlation coefficient, generalizing the Kendall tau. Moreover, τ_{FG} being a closed form transport map we use coordinate-wise cost functions to quantify contrasts between F and G. Clustering, quantization and classification follow for sampled bivariate distributions, as well as multiple tests based on the joint contrasts between F and a collection of G's of reference.

Notice that estimating curves and distributions along curves from empirical probabilities of quadrants is far from being obvious from the theoretical wiewpoint. However we establish in [3] sharp asymptotics and non asymptotic Gaussian field approximations of the Kendall geometry of F. Non standard uniform central limit theorems (CLT's) for $\mathcal{G}_{F,n}$ and $\tau_{n,m}$ are obtained through empirical processes and Brownian coupling techniques, thanks to the exact formulas and geometrical intuition we develop below. As a consequence, most of the statistical tools derived from $\mathcal{G}_F$ and τ_{FG} at Sections 5.5 and 6 could be controlled by explicit CLT's and p-values tabulated by simulations. This is beyond the scope of this exploratory paper that focuses on definitions, basic properties and visual numerical experiments.

1.4 Comments

Let put forward a few important facts that will become obvious along the paper.

First, the generator $\mathcal{G}_F$ and the quantile transform map τ_{FG}, are basis dependent. As a matter of fact the natural basis in statistics is provided by the marginals of the original $r.v.$ determining F. However one can define basis independent contrasts and tests by minimizing among the rotations of the basis, or averaging the values computed over all or some rotations. Likewise for relevant statistical features such as contours and depth, and this is what we actually illustrate.

Second, $\mathcal{G}_{F,n}$ and $\tau_{n,m}$ are well defined whatever the samples, whereas $\mathcal{G}_F$ and τ_{FG} are defined for F and G supported by finite or infinite parallel rectangles, for sake of simplicity – including the important case of copulas. This could be extended to F and G supported by a convex or a single connex component having smooth enough boundary.

Third, the empirical quantile transform $\tau_{n,m}$ can be made almost everywhere continuous, and differs from a discrete transport plans of one sample to the other. Indeed, the new paradigm at work when using $\tau_{n,m}$ is to estimate the curves of F and G separately and non parametrically then match them.

In particular, the empirical couplings $(\mathcal{G}_{F,n}(Z_j), \mathcal{G}_{G,m}(Z_j), j = 1, ..., k$, generate a fully rank-correlated sample with marginals almost F and G. Such a new sample has no repetitions since the empirical generator $\mathcal{G}_{F,n}(Z)$ produces realizations that are not sample points. The variability is increased when $\mathcal{G}_{F,n}$ is used to bootstrap instead of F_n. This may have some computational advantages for Monte-Carlo type resampling methods to derive critical values for tests.

Lastly, the singular fact that the univariate generator $r.v.$ Z is a special case of the generated $r.v.$'s through the identity mapping is misleading in higher dimension where $F(X)$ strongly depends on F. It is noteworthy that $\mathcal{G}_F$ of Theorem 1 is never the identity, even for the uniform distribution. The rank $\mathcal{G}_F^{-1}$ is universal since $\mathcal{G}_F^{-1}(X)$ is always uniform, for $F \in \overline{\mathbb{F}}$. The information catched by estimating $\mathcal{G}_F^{-1}$ is mainly designed to built new geometrical data analysis tools.

1.5 Some other approaches

As no natural ordering shows up in higher dimension, most generalizations of quantiles are based on various notions of quantile sets, not quantile points or generators. In [7] and [15] quantiles are sets selected among a predetermined collection, by using a minimum volume or differential gradient criterium, which leads to an explicit M-estimator asymptotic theory. In the above spirit we prefer quantile shapes entirely determined by F instead of selected among a predetermined class. Other kind of nested increasing sets have been proposed. Alternatives combine univariate quantiles of marginals, or use directional quantiles through various projections. A rather popular approach is based on quantile regression, hyperplanes and quantile contours – see for instance [4], [16] and [13]. In [1] directional quantiles define surfaces seen from any observation point, leading to uniform limit theorems with dimension free rates.

In more recent approaches such as [5] and [10], quantile points and rank vectors are determined by transporting F to an arbitrary distribution of reference. This makes the univariate optimal coupling aspect a motivating definition of the rank to quantile transform on $\mathbb{R}^d$ and provides a universal generator that do not require orientation. In dimension $d \leq 4$ the estimation of transport costs has asymptotic guaranties – see [19] and [18]. However the optimal transport maps used are implicit and thus enjoy few properties to exploit from the statistical viewpoint. They are numerically approximated, thus asking for small samples, and no asymptotic theory is available. The main difference with our explicit map τ_{FG} is that the role of the easily estimated F is naturally lost.

1.6 Overview

The paper is organized as follows. In Section 2 we formally define the bivariate rank and quantile points based on the generator and the associated geometry. In Section 3 we give some properties of the obtained quantile transform τ_{FG} related to some partial optimality. At Section 4 we derive a few statistical properties of Kendall ranks. Section 5 is devoted to the algorithmic definition of the empirical versions $\mathcal{G}_{F,n}$ and $\tau_{n,m}$ of the generator and quantile transform map, with numerical examples. In Section 6 we propose new statistical quantities and geometrical features based on Kendall quantiles, ranks and transform.

2 Generator and quantile transform map

In this section we define the generator $\mathcal{G}_F$ of $F \in \mathbb{F}$ with probability measure P, and the quantile transform (4). Then we derive Theorem 1 for bounded rectangles. Extension to $\mathbb{R}^2$ then easily follows.

2.1 Generator equation along Q-curves

For $F \in \mathbb{F}$ let denote F_1 and F_2 the $d.f.'s$ on $\mathbb{R}$ of the marginal $r.v.$ $X_1 = \langle X, e_1 \rangle$ and $X_2 = \langle X, e_2 \rangle$. For $\alpha \in [0,1]$ consider the α-level set of F,

$$\mathrm{Q}_F(\alpha) = \left\{ x \in \overline{\mathcal{R}}_X : F(x) = \alpha \right\} \tag{5}$$

that we call the α-th Q-curve of F, joining the point $(F_1^{-1}(\alpha), x_2^+)$ to $(x_1^+, F_2^{-1}(\alpha))$. Hence $\mathrm{Q}_F(1) = \left\{ (x_1^+, x_2^+) \right\}$ is the upper-right corner and $\mathrm{Q}_F(0) = \left\{ (x_1, x_2) : x_1 = x_1^- \text{ or } x_2 = x_2^- \right\}$ is the lower left half-perimeter. Define $\mathbb{Q}_F(0) = \mathrm{Q}_F(0)$, $\mathbb{Q}_F(1) = \overline{\mathcal{R}}_X$ and, for $\alpha \in [0,1]$, the α-th Q-set of F

$$\mathbb{Q}_F(\alpha) = \left\{ x \in \overline{\mathcal{R}}_X : F(x) \leq \alpha \right\} = \bigcup\nolimits_{a=0}^{\alpha} \mathrm{Q}_F(a). \tag{6}$$

Definition 2. *Let the Kendall $d.f.$ of F be*

$$K_F(\alpha) = P(\mathbb{Q}_F(\alpha)) = \mathbb{P}(F(X) \leq \alpha), \quad \alpha \in [0,1]\,. \tag{7}$$

In other words K_F is the $d.f.$ of the $r.v.$ $F(X)$. We first notice that K_F only depends on the copula function of $F \in \mathbb{F}$. Since $F \in \mathbb{F}$ has positive density on the open rectangle $\mathcal{R}_X$, we have the following result from [23]:

Proposition 3. *If $(F, G) \in \mathbb{F} \times \mathbb{F}$ have same copula then $K_F = K_G$.*

Restricting F to a parametric or semi-parametric family may allow K_F to identify F. For instance, if F is an Archimedean copula, that is $F(x) = \phi^{-1}(\phi(x_1) + \phi(x_2))$ then $K_F(\alpha) = \alpha - \phi(\alpha)/\phi'(\alpha)$ characterizes F.

We shall also use the fact that for $F \in \mathbb{F}$, K_F is $\mathcal{C}^2$.

Proposition 4. *If F is $\mathcal{C}^2$ and has continuous positive density f on the open rectangle $\mathcal{R}_X$ then K_F has continuous positive density on $(0,1)$, the collection of Q-curves $\alpha \to Q_F(\alpha)$ determine F and the collection of Q-sets $\alpha \to \mathbb{Q}_F(\alpha)$ determine F. If moreover f is $\mathcal{C}^1$ on $\mathcal{R}_X$ then K_F is $\mathcal{C}^2$.*

Before proving Proposition 4 we need to describe more precisely (5) as a parametric curve $\mathrm{Q}_F(\alpha, t)$ with $t \geq 0$. Let ∇F denote the gradient of F and $\overline{\nabla} F$ its right-oriented orthogonal vector tangent to the smooth curve $\mathrm{Q}_F(\alpha)$. Thus ∇F and $\overline{\nabla} F$ satisfy $\langle \nabla F, e_1 \rangle \geq 0$, $\langle \nabla F, e_2 \rangle \geq 0$, $\langle \overline{\nabla} F, e_1 \rangle \geq 0$ and $\langle \overline{\nabla} F, e_2 \rangle \leq 0$.

Definition 5. *For any $\alpha \in (0,1)$ define $t \in \mathbb{R}^+ \to Q_F(\alpha, t) \in Q_F(\alpha)$ to be the solution of the ordinary differential equation*

$$\frac{dQ_F(\alpha, t)}{dt} = \overline{\nabla} F(Q_F(\alpha, t)), \quad Q_F(\alpha, 0) = (F_1^{-1}(\alpha), x_2^+). \tag{8}$$

Write $t_F : x \in \mathcal{R}_X \to t_F(x)$ the unique solution (in t) of $Q_F(F(x), t) = x$. The total time along $Q_F(\alpha)$ is

$$T_F(\alpha) = t_F((x_1^+, F_2^{-1}(\alpha))) = \min \left\{ t : Q_F(\alpha, t) = (x_1^+, F_2^{-1}(\alpha)) \right\}. \tag{9}$$

Define the mass-time density on $[0, T_F(\alpha)]$ *to be*

$$f_\alpha(t) = \frac{f(Q_F(\alpha, t))}{k_F(\alpha)}, \tag{10}$$

where

$$k_F(\alpha) = \int_0^{T_F(\alpha)} f(Q_F(\alpha, t))dt. \tag{11}$$

Proof of Proposition 4. It holds $\|\overline{\nabla} F\|_2 = \|\nabla F\|_2 > 0$ on $\mathcal{R}_X$, as $f > 0$. Write f_1 the density of F_1, and observe that $\|\overline{\nabla} F(\mathrm{Q}_F(\alpha, 0))\|_2 > 0$ for $\alpha \in (0, 1)$, since $f_1(F_1^{-1}(\alpha)) > 0$. The existence of k_F follows from a change of variable onto the parametrization of Definition 5. Recall that F is $\mathcal{C}^2$ and $\overline{\nabla} F$ is $\mathcal{C}^1$ thus Q_F is $\mathcal{C}^1$. Since $\left(\nabla F/\|\nabla F\|_2, \overline{\nabla} F/\|\nabla F\|_2\right)$ is an orthonormal basis and $F(\mathrm{Q}_F(\alpha, t)) = \alpha$ we get, by (8),

$$\frac{d\mathrm{Q}_F(\alpha, t)}{d\alpha}.\nabla F(\mathrm{Q}_F(\alpha, t)) = 1,$$
$$\left| det\left(\frac{d\mathrm{Q}_F(\alpha, t)}{dt}, \frac{d\mathrm{Q}_F(\alpha, t)}{d\alpha}\right)\right| = 1.$$

Therefore, by putting $x = \mathrm{Q}_F(\alpha, t)$,

$$K_F(\alpha) = \int_{x \in \mathbb{Q}_F(\alpha)} f(x)dx = \int_0^\alpha \int_0^{T_F(a)} f(\mathrm{Q}_F(a, t))dtda = \int_0^\alpha k_F(a)da.$$

Hence k_F is a positive density of K_F on $(0, 1)$. Assuming that f is $\mathcal{C}^1$, the alternative representation (14) below shows that k_F is $\mathcal{C}^1$ and, by (11), $T_F(\alpha)$ is differentiable. Next, for $\alpha' > \alpha$ it holds $\mathrm{Q}_F(\alpha) \cap \mathrm{Q}_F(\alpha') = \emptyset$ and $\mathbb{Q}_F(\alpha) \subset \mathbb{Q}_F(\alpha')$ thus the Q-curves determine $F(x) = \min\{q : x \in \mathrm{Q}_F(q)\}$ and the Q-sets determine $F(x) = \min\{q : x \in \mathbb{Q}_F(q)\}$ for $x \in \mathcal{R}_X$. □

2.2 Generator property

In order to define the generator we need to change coordinate in the solution $\mathrm{Q}_F(\alpha, t)$ of (8).

Definition 6. *For* $\alpha \in (0, 1)$ *let the mass-time d.f. along the* α*-th Q-curve of* F *be*

$$F_\alpha(t) = \int_0^t f_\alpha(s)ds \in (0, 1), \quad \textit{for } t \in [0, T_F(\alpha)]\,. \tag{12}$$

Write $F_\alpha^{-1} : [0, 1] \to [0, T_F(\alpha)]$ *its inverse function and set*

$$x_F(\alpha, u) = Q_F(\alpha, t) \textit{ such that } F_\alpha(t) = u, \quad \textit{for } u \in [0, 1]\,. \tag{13}$$

Thus $u \to x_F(\alpha, u) = \mathrm{Q}_F(\alpha, F_\alpha^{-1}(u))$ is a parametrization of $\mathrm{Q}_F(\alpha)$ by $[0, 1]$ through the Q-curve conditional probability. Likewise, $x_F(\alpha, u)$ is a parametrization of $\mathcal{R}_X$ by $\mathbb{U} = (0, 1)^2$ through the geometry.

Remark 7. *Combined with (2), (13) induces a strict order on* $\mathcal{R}_X$ *that characterizes* F*. Namely* $x_F(\alpha, u) < x_F(\alpha^+, u^+)$ *if* $(\alpha, u) < (\alpha^+, u^+)$ *in the sense of (2), which is equivalent to (3). The idea behind is that the mass has been ordered to prepare its generation then its transportation.*

At this stage, $(\alpha, u) \in (0, 1)^2$ stands as a generator coordinate system and $x_F(\alpha, u)$ from (13) as a generator mapping process for F. Starting from two independent uniform $r.v.$'s, x_F can be used as follows to perfectly simulate an X with distribution F.

Theorem 8. *Let* $F \in \mathbb{F}$ *and* $Z = (Z_1, Z_2)$ *be a uniform* $r.v.$ *on* $\mathbb{U}$*. Then the* $r.v.$ $x_F(K_F^{-1}(Z_1), Z_2)$ *has distribution* F*.*

Proof. First we show that the one to one mapping $x_F(\alpha, u)$ is also $\mathcal{C}^1$ from $\mathbb{U}$ to $\mathcal{R}_X$, and thus is a $\mathcal{C}^1$ diffeomorphism. The smoothness with respect to u is clear, what remains to prove is the smoothness of x_F with respect to α. From $x_F(\alpha, u) = \mathrm{Q}_F(\alpha, F_\alpha^{-1}(u))$ we only need to check the continuous differentiability of $k_F(\alpha)$. Recall that F is $\mathcal{C}^2$. Let parametrize $\mathrm{Q}_F(\alpha)$ with $x_1 \in \left[F_1^{-1}(\alpha), x_1^+\right]$. There exists r_α such that $x_2 = r_\alpha(x_1)$ when $x \in \mathrm{Q}_F(\alpha)$. As $F(x_1, r_\alpha(x_1)) = \alpha$ it holds

$$\frac{\partial r_\alpha(x_1)}{\partial \alpha} = -\frac{1}{\frac{\partial F(x_1, r_\alpha(x_1))}{\partial x_2}}.$$

Now, the fact that F is $\mathcal{C}^2$ readily implies that $\partial r_\alpha(x_1)/\partial\alpha$ is $\mathcal{C}^1$ with respect to (α, t), then integrating over $\mathbb{Q}_F(\alpha)$ of (6) yields

$$K_F(\alpha) = \alpha + \int_{F_1^{-1}(\alpha)}^{x_1^+} dx_1 \int_0^{r_\alpha(x_1)} f(x_1, x_2)dx_2$$

then deriving with respect to α gives

$$k_F(\alpha) = 1 - \frac{1}{f_1(F_1^{-1}(\alpha))}\int_{x_2^-}^{x_2^+} f(F_1^{-1}(\alpha), x_2)dx_2 + \int_{F_1^{-1}(\alpha)}^{x_1^+} \frac{\partial r_\alpha(x_1)}{\partial\alpha} f(x_1, r_\alpha(x_1))dx_1. \tag{14}$$

From (14) we get, as f and $\partial r_\alpha(x_1)/\partial\alpha$ are $\mathcal{C}^1$ and $\partial f/\partial x_1$ is continuous, that $k_F(\alpha)$ is $\mathcal{C}^1$. Next we prove that the function defined on $\mathbb{U}$ by $H(z_1, z_2) = x_F(K_F^{-1}(z_1), z_2)$ satisfies $|det\ \nabla H(z_1, z_2)| = \frac{1}{f(H(z_1,z_2))}$. Since $\partial \mathrm{Q}_F(\alpha, s)/\partial t = \overline{\nabla} F(\mathrm{Q}_F(\alpha, t))$ and $\partial F_\alpha(t)/\partial t = f_\alpha(t)$, we have

$$\frac{\partial H(z_1, z_2)}{\partial z_2} = \overline{\nabla} F\left(x_F(K_F^{-1}(z_1), z_2)\right) \frac{k_F(K_F^{-1}(z_1))}{f(x_F(K_F^{-1}(z_1), z_2))} \tag{15}$$

then it holds $F(H(z_1, z_2)) = K_F^{-1}(z_1)$ and thus

$$\frac{\partial H(z_1, z_2)}{\partial z_1}\nabla F(H(z_1, z_2)) = \frac{1}{k_F(K_F^{-1}(z_1))}. \tag{16}$$

The proof is complete, $\left(\nabla F/\|\nabla F\|_2, \overline{\nabla} F/\|\nabla F\|_2\right)$ being orthonormal. □

2.3 Generator map

We are now ready to define the class of generators of $r.v.$ with distribution in $\mathbb{F}$ mentioned in Theorem 1. Remind (7), (12) and (13).

Definition 9. *The generator $\mathcal{G}_F : \mathbb{U} \to \mathcal{R}_X$ is defined to be*

$$\mathcal{G}_F(z) = Q_F\left(K_F^{-1}(z_1), F^{-1}_{K_F^{-1}(z_1)}(z_2)\right) = x_F(K_F^{-1}(z_1), z_2), \quad z \in \mathbb{U}.$$

The universal generator map is $\mathcal{G} : F \in \mathbb{F} \to \mathcal{G}(F) = \mathcal{G}_F$.

By Theorem 8 for $Z \sim U$, the $r.v.$ $\mathcal{G}_F(Z)$ has distribution F. By the proof of Theorem 8, and Proposition 4, $\mathcal{G}_F$ is a $\mathcal{C}^1$ diffeomorphism from $\mathbb{U}$ to $\mathcal{R}_X$. For $F \in \mathbb{F}$, $\mathcal{G}_F$ can not preserve vertical lines, which excludes the identity.

To characterize $\mathcal{G}_F$ by the Kendall geometry of F, we need to properly define the R-curves :

Definition 10. *Given $z_2 \in (0, 1)$, the z_2-th R-curve of $F \in \mathbb{F}$ is the parametrized curve $R_F(z_2) : \alpha \in (0, 1) \to x_F(\alpha, z_2)$. Then we also define the z_2-th R-set as*

$$\mathbb{R}_F(z_2) = \bigcup_{z_2'=0}^{z_2} R_F(z_2'). \tag{17}$$

Thus the geometrical definition of $\mathcal{G}_F(z)$ is the unique intersection of the $K_F^{-1}(z_1)$-th Q-curve and the z_2-th R-curve. For short we call the R and Q curves intersecting at $\mathcal{G}_F(z)$ the z_1 and z_2 curves.

Remark 11. *In the generator geometry $\mathcal{G}_F(z_1, z_2)$ the parameter z_1 is the probability "below the Q-curve" and the parameter z_2 is the probability "along the Q-curve". We can also say that z_2 is the probability "above the R-curve", or "of the R-set", and z_1 the probability "of the Q-set". The Q and R curves are of equal importance to geometricaly locate $X = \mathcal{G}_F(Z)$ in $\mathcal{R}_X$. However the R-curves alone are not enough to determine the distribution. They can be deduced from the Q-curves – that determine F – as exploited by the algorithm of Section 5.*

Lets give a basic example that is not the simplest since K_F is not trivially invertible.

Example 12. *Let us build $\mathcal{G}_F$ explicitly for $F = U$, to show that $\mathcal{G}_U$ is not at all the identity. We easily get $Q_U(\alpha) = \{x : x_1x_2 = \alpha\}$, $K_U(\alpha) = \alpha - \alpha\log\alpha$, $k_F(\alpha) = T_U(\alpha) = -\log\alpha$, $Q_U(\alpha, t) = (\alpha e^t, e^{-t})$, and $U_\alpha(t) = -t/\log\alpha$, $0 \le t \le T_U(\alpha)$. For $(z_2, \alpha) \in (0, 1)^2$, $R_U(z_2)$ is the (power) curve $(\alpha^{1-z_2}, \alpha^{z_2})$. Then, by Definition 9,*

$$\mathcal{G}_U(Z) = \left(K_U^{-1}(Z_1)^{(1-Z_2)}, K_U^{-1}(Z_1)^{Z_2}\right).$$

After some computations one can verify that $\mathcal{G}_U(Z)$ actually has same distribution U as Z, but $\mathcal{G}_U(z) \neq z$ for $z \in \mathbb{U}$.

2.4 Quantile transform map

The main property making the coordinate system of $\mathcal{G}$ universal is that all couplings $(\mathcal{G}_F(Z), \mathcal{G}_G(Z))$ are simultaneously ordered in the sense of (3). Write $\mathcal{T}_1(F,G)$ the set of continuously differentiable transport maps from F to G. Hence $\tau \in \mathcal{T}_1(F,G)$ is one to one and $\tau(X)$ has $d.f.$ G.

Proposition 13. *For $(F,G) \in \mathbb{F} \times \mathbb{F}$, the map $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$ satisfies $\tau_{FG} \in \mathcal{T}_1(F,G)$.*

By combining Theorem 8, Definition 9 and Proposition 13 we have proved Theorem 1. Clearly, τ_{FG} is the unique one to one mapping between $\mathcal{R}_X$ and $\mathcal{R}_Y$ preserving the Kendall ordering of F and G.

3 Probabilistic properties of the quantile transform τ_{FG}

3.1 Q-curves optimal transport

In this section we generalize the usual formula of the optimal transport between two distributions on $\mathbb{R}$ to the case of two uniform distributions U_x and U_y on any smooth curves x and y in $\mathbb{R}^d$ that are globally coordinate-wise co-monotonic. The above Q-curves are a special case, for $d = 2$.

Theorem 14. *Fix $d > 1$. Let $x(t)$ and $y(t)$ be two $\mathcal{C}^1$ curves in $\mathbb{R}^d$ parametrized by $t \in (0,1)$. For $1 \leq i \leq d$, denote $x_i(t)$ and $y_i(t)$ their coordinates and assume that for any $(t_1,t_2) \in (0,1)^2$ their derivatives satisfy $x_i'(t_1)y_i'(t_2) > 0$. Let c be any cost of the form*

$$c(x,y) = \sum_{i=1}^{d} c_i(x_i - y_i)$$

where the function c_i are $\mathcal{C}^1$ on $\mathbb{R}$, $\mathcal{C}^2$ on $\mathbb{R}^*$, strictly convex, non negative, null at 0 and satisfy, for $1 \leq i \leq d$ and all x_i, x_i', y_i, y_i' in $\mathbb{R}$,

$$-\int_{x_i}^{x_i'}\int_{y_i}^{y_i'} c_i''(x-y)dxdy = c_i(x_i' - y_i') - c_i(x_i' - y_i) - c_i(x_i - y_i') + c_i(x_i - y_i). \tag{18}$$

Consider two uniform $r.v.$ on $(0,1)$, U and V. Then the c-optimal transport map between $X = x(U)$ and $Y = y(V)$ is given by $(x(U), y(U))$.

Proof. We mimic the proof on the real line, coordinate by coordinate. Let us estimate $\mathbb{E}(c(x(U) - y(V)))$ whatever the copula π, that is the joint distribution of (U,V). First observe that, for $1 \leq i \leq d$,

$$\begin{aligned} c_i(x_i(u) - y_i(v)) = {} & c_i(x_i(u) - y_i(0)) + c_i(x_i(0) - y_i(v)) - c_i(x_i(0) - y_i(0)) \\ & - \int_0^u \int_0^v c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t)dsdt. \end{aligned}$$

Since $\mathbb{E}(c_i(x_i(U) - y_i(0)))$ and $\mathbb{E}(c_i(x_i(0) - y_i(V)))$ only depend on the known marginals $x_i(U)$ and $y_i(V)$ we only need to evaluate $\mathbb{E}(-\int_0^U \int_0^V c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t)dsdt)$. By denoting Π the distribution function of π it holds, applying Fubini's theorem,

$$\begin{aligned} & \mathbb{E}\left(-\int_0^U \int_0^V c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t)dsdt\right) \\ & = -\int_0^1 \int_0^1 \int_s^1 \int_t^1 \pi(du,dv)c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t)dsdt \\ & = -\int_0^1 \int_0^1 c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t)\left(1 - s - t + \Pi(s,t)\right)dsdt. \end{aligned}$$

Here $c_i''\left(x_i(s) - y_i(t)\right)x_i'(s)y_i'(t) \geqslant 0$ by hypothesis, and it is well known that the maximum of $\Pi(s,t)$ is achieved for $\Pi^+(s,t) = s \wedge t$, that is the distribution of (U,U). □

Write $p_x = x \circ U$ and $p_y = y \circ U$ the distribution on the parametrized curves x and y associated to a uniform parameter. Theorem 14 yields

$$\begin{aligned} W_c(p_x,p_y) & = \min_{X \sim p_x, Y \sim p_y} \mathbb{E}\left(c(X,Y)\right) \\ & = \int_0^1 c(x(u),y(u))du = \sum_{i=1}^{d} \int_0^1 c_i(x_i(u) - y_i(u))du. \end{aligned}$$

Example 15. *All the p-norms, $p > 1$, satisfy (18). One can also use various weighted sums of power functions, namely, for $p_i > 1$ and $a_i > 0$,*

$$c(x) = \sum_{i=1}^{d} a_i |x_i|^{p_i}.$$

Remark 16. *Actually the condition (18) can be weakened but this is not the point here. For sake of simplicity we use coordinate-wise costs, however the result is valid for costs and curves satisfying, for all $(u, v) \in (0, 1)^2$,*

$$\int_0^u \int_0^v x'(s)^T c''\left(x(s) - y(t)\right) y'(t) ds dt + c(x(u) - y(v)) = \phi(x, u) + \psi(y, v)$$

and $x'(s)^T c''\left(x(s) - y(t)\right) y'(t) \geq 0$ for all $(s, t) \in \mathbb{U}$, with c'' the Hessian matrix of c and $x'(s)^T$ the transposed gradient of x at time t.

Let denote $\tilde{F}_\alpha$ the probability measure on $\mathrm{Q}_F(\alpha)$ of (5) putting measure $F_\alpha(t_1) - F_\alpha(t_0)$ from (12) to the Q-curve arc joining $\mathrm{Q}_F(\alpha, t_0)$ to $\mathrm{Q}_F(\alpha, t_1)$, for any $0 < t_0 < t_1 < T_F(\alpha)$ of (9). Consider again $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$. Let c_1 and c_2 be $\mathcal{C}^1$, strictly convex functions on $\mathbb{R}$, $\mathcal{C}^2$ and positive on $\mathbb{R}^*$, such that $c_1(0) = c_2(0) = 0$ and (18). Consider $c(x, y) = c_1(x_1 - y_1) + c_2(x_2 - y_2)$ for $x = (x_1, x_2)$, $y = (y_1, y_2)$. This includes all Wasserstein costs W_p, for $c_1(w) = c_2(w) = |w|^p$ and $p > 1$.

Corollary 17. *For any $\alpha \in (0, 1)$, $(F, G) \in \mathbb{F} \times \mathbb{F}$, τ_{FG} c-optimally transports $\tilde{F}_\alpha$ onto $\tilde{G}_{K_G^{-1} \circ K_F(\alpha)}$.*

Proof. We can straightforwardly apply Theorem 14 for $d = 2$ to the Q-curves parametrized by $u \in (0, 1)$, namely $x(u) := x_F(\alpha, u) = \mathrm{Q}_F(\alpha, F_\alpha^{-1}(u))$ and $y(u) := x_G(\beta, u) = \mathrm{Q}_G(\beta, G_\beta^{-1}(u))$. Clearly, whatever $(\alpha, \beta) \in (0, 1)^2$ and $(F, G) \in \mathbb{F} \times \mathbb{F}$, the two components of $\overline{\nabla} F$ and $\overline{\nabla} G$ have always the same sign, hence the assumptions of Theorem 14 are satisfied. □

3.2 R-curves optimal transport

Let show that τ_{FG} satisfies a property similar to Corollary 17 for the R-curves. Denote $\tilde{K}_{F,z_2}$ the probability measure on $\mathrm{R}_F(z_2)$ putting measure $K_F(\alpha_1) - K_F(\alpha_0)$ to the R-curve arc joining $x_F(\alpha_0, z_2)$ to $x_F(\alpha_1, z_2)$ along $\mathrm{R}_F(z_2)$, for any $0 < \alpha_0 < \alpha_1 < 1$. For $G \in \mathbb{F}$ consider $\mathrm{R}_G(z_2)$ and the probability measure $\tilde{K}_{G,z_2}$. Observe that $\mathrm{R}_G(z_2)$ is also the image curve

$$\tau_{FG}(\mathrm{R}_F(z_2)) : \alpha \in (0, 1) \to x_G(K_G^{-1} \circ K_F(\alpha), z_2),$$

however with a different parametrization. By Definitions 2 and 6,

$$F((0, \alpha) \times (1 - x_F((0, \alpha), z_2))) = z_2 K_F(\alpha),$$
$$G((0, \beta) \times (1 - x_G((0, \beta), z_2))) = z_2 K_G(\beta),$$

and when $\beta = K_G^{-1} \circ K_F(\alpha)$ it holds $K_G(\beta) = K_F(\alpha)$. This simply means that the R-curves $\mathrm{R}_F(z_2)$ divide the open square in two open subsets of probability z_2 and $1 - z_2$ and of probability $z_2 K_F(\alpha)$ and $(1 - z_2) K_F(\alpha)$ when stopped at "mass-time" α. The same holds for $\mathrm{R}_G(z_2)$ and K_G.

As observed above, τ_{FG} maps $\mathrm{R}_F(z_2)$ onto $\mathrm{R}_G(z_2)$ and simply consists in transporting K_F onto K_G by $\alpha \to K_G^{-1} \circ K_F(\alpha)$, which actually is on $(0, 1)$ the optimal transport of K_F to K_G. The map τ_{FG} always preserves R-curves and also optimally transports them if the supports are equal.

Proposition 18. *For any $\alpha \in (0, 1)$ and $(F, G) \in \mathbb{F} \times \mathbb{F}$ such that $\mathcal{R}_X = \mathcal{R}_Y$, τ_{FG} c-optimally transports $\tilde{K}_{F,z_2}$ onto $\tilde{K}_{G,z_2}$.*

Proof. As a mater of fact there is only two $\mathcal{C}^1$ smooth transports and the restriction of τ_{FG} is optimal. Observe that if a transport of $\tilde{K}_F$ on $R_F(z_2)$ to $\tilde{K}_G$ on $R_G(z_2)$ is $\mathcal{C}^1$ and given by the change of index $l(\alpha)$ then

$$k_F(\alpha) = g(x_G(l(\alpha), z_2)) \left| \frac{dx_G(l(\alpha), z_2)}{d\alpha} l'(\alpha) \right| = k_G(l(\alpha))) |l'(\alpha)|.$$

Hence $l'(\alpha) = \pm \dfrac{k_F(\alpha)}{k_G(l(\alpha)))}$. The choice "+" implies $l(\alpha) = K_G^{-1} \circ K_F(\alpha)$ since $l(0) = 0$ and $l(1) = 1$. Now, the curves $\mathrm{R}_F(z_2)$ and $\mathrm{R}_G(z_2)$ having the same limiting points $\lim_{\alpha \to 1} x_F(\alpha, z_2) = (x_1^+, x_2^+)$ and $\lim_{\alpha \to 0} x_F(\alpha, z_2) = (x_1^-, x_2^-)$, the choice "−" is obviously not better.

□

Remark 19. *In other words, the parametrization of the curves $R_F(z_2)$ and $R_G(z_2')$ by α gives the $\mathcal{C}^1$ optimal transport maps of $\tilde{K}_{F,z_2}$ onto $\tilde{K}_{G,z_2'}$. This remains true when $\mathcal{R}_X \neq \mathcal{R}_Y$ – without elegant proof.*

At this stage we have defined the geometry characterizing F as the system of Q-curves and R-curves driven by the new coordinates (α, z_2) or (z_1, z_2). These curves are optimally send by τ_{FG} on the analog geometry of G with respect to a family of costs including all the p-norm costs, $p > 1$.

3.3 Same copula case

It is known ([6]) that F and G share the same copula if, and only if, the optimal transport map of F on G is the product of the optimal transport maps coordinate by coordinate. Interestingly, in this case the quantile transform map τ_{FG} coincides with this product, which is a nice consequence of using the $d.f.$ level sets – yet not obvious at first look.

Proposition 20. *If $(F, G) \in \mathbb{F} \times \mathbb{F}$ have the same copula then $\tau_{FG}(x) = (\tau_{F_1G_1}(x_1), \tau_{F_2G_2}(x_2))$ is the product of the quantile transform of the marginals of F onto those of G.*

Proof. Write the common copula $C_F = C_G$. The transport map is

$$\tau_{FG}(\mathrm{Q}_F(K_F^{-1}(z_1), F^{-1}_{K_F^{-1}(z_1)}(z_2))) = \mathrm{Q}_G(K_G^{-1}(z_1), G^{-1}_{K_G^{-1}(z_1)}(z_2)).$$

Write $\overline{F} = (F_1, F_2)$ and $\overline{G}^{-1} = (G_1^{-1}, G_2^{-1})$ for convenience. Let show that for any fixed z_1 the two following curves are the same,

$$z_2 \to \mathrm{Q}_G(K_G^{-1}(z_1), G^{-1}_{K_G^{-1}(z_1)}(z_2)),$$
$$z_2 \to \overline{G}^{-1} \circ \overline{F}(\mathrm{Q}_F(K_F^{-1}(z_1), F^{-1}_{K_F^{-1}(z_1)}(z_2))).$$

By Proposition 3 we have $K_F = K_G$ and, for $z_2 = 0$ it holds

$$\mathrm{Q}_G(K_G^{-1}(z_1), G^{-1}_{K_G^{-1}(z_1)}(0)) = (G_1^{-1}(K_G^{-1}(z_1)), 1) = \overline{G}^{-1} \circ \overline{F}(F_1^{-1}(K_F^{-1}(z_1)), 1)$$

hence they start from the same point. It is then sufficient to verify that they are driven by the same differential equation with respect to z_2. Put $\alpha = K_F^{-1}(z_1) = K_G^{-1}(z_1)$. Let use the shortcut $g_i(.)$ for the value of the density g_i at the coordinate i of $\mathrm{Q}_G(\alpha, G_\alpha^{-1}(z_2))$ and mutatis mutandis for f_i and Q_F and denote $.*$ the coordinate-wise product. Then

$$\begin{aligned}\frac{d\mathrm{Q}_G(\alpha, G_\alpha^{-1}(z_2))}{dz_2} &= \frac{1}{g_\alpha(G_\alpha^{-1}(z_2))}(g_2(.), g_1(.))^T .* \overline{\nabla} C_G(.,.) \\ &= \frac{k_G(\alpha)}{c_G(\overline{G}(.,.))g_1(.)g_2(.)}(g_2(.), g_1(.))^T .* \overline{\nabla} C_G(.,.)\end{aligned}$$

with $c_F = c_G$ the copula density, and

$$\begin{aligned}\frac{d\,\overline{G}^{-1} \circ \overline{F}(\mathrm{Q}_F(\alpha, F_\alpha^{-1}(z_2)))}{dz_2} &= \left(\frac{f_1(.)}{g_1(.)}, \frac{f_2(.)}{g_2(.)}\right)^T .* \frac{1}{f_\alpha(F_\alpha^{-1}(z_2))}(f_2(.), f_1(.))^T .* \overline{\nabla} C_F(.,.) \\ &= \left(\frac{f_1(.)}{g_1(.)}, \frac{f_2(.)}{g_2(.)}\right)^T .* \frac{k_F(\alpha)}{c_F(\overline{F}(.,.))f_1(.)f_2(.)}(f_2(.), f_1(.))^T .* \overline{\nabla} C_F(.,.).\end{aligned}$$

Inspecting each component of the two formulas reveals the same differential equation. Thus τ_{FG} is the product of the marginals quantile transforms. □

3.4 Extension to $\overline{\mathbb{F}}$

First we consider the extension to whole $\mathbb{R}^2$. Let $\bar{\mathcal{F}}$ be the set of the $\mathcal{C}^2$ distribution functions F on $\mathbb{R}^2$ with $\mathcal{C}^1$ positive density f. The definitions of the Q-sets, R-sets (5), Q-curves, R-curves (6) and Kendall distribution (7) remain the same, however the Q-curves, R-curves and the Q-sets, R-sets are no more bounded. The choice of each x_α below is left arbitrary since it leads to the same time-mass geometry of F.

Definition 21. *For any $\alpha \in (0, 1)$, any $x_\alpha \in Q_F(\alpha)$ define the α-th Q-curve $Q_F(\alpha, x_\alpha, .)$, indexed by time $t \in \mathbb{R}$, to be the solution of the ordinary differential equation*

$$\frac{dQ_F(\alpha, x_\alpha, t)}{dt} = \overline{\nabla} F(Q_F(\alpha, x_\alpha, t)), \quad Q_F(\alpha, x_\alpha, 0) = x_\alpha. \tag{19}$$

We have $Q_F(\alpha) = \{Q_F(\alpha, x_\alpha, t), t \in \mathbb{R}\}$. *Define* F_α *to be the time distribution on* $\mathbb{R}$ *with density*

$$f_\alpha(t) = \frac{f(Q_F(\alpha, x_\alpha, t))}{k_F(\alpha)},$$

where

$$k_F(\alpha) = \int_{\mathbb{R}} f(Q_F(\alpha, x_\alpha, t))dt.$$

We have $k_F(\alpha) < \infty$ and $k_F(\alpha)$ does not depend on $x_\alpha \in \mathrm{Q}_F(\alpha)$. Indeed, for any $x_1 \in \mathrm{Q}_F(\alpha)$ there exists a unique $t_1 \in \mathbb{R}$ depending on x_α such that $x_1 = \mathrm{Q}_F(\alpha, x_\alpha, t_1)$ and hence $\mathrm{Q}_F(\alpha, x_\alpha, t - t_1) = \mathrm{Q}_F(\alpha, x_1, t)$. The time-mass at x, is $t_F(x)$ such that $x = \mathrm{Q}_F(F(x), x_{F(x)}, t_F(x))$. The local $d.f.$ along the Q_F-curves and Q_G-curves are

$$F_\alpha(t) = \int_{-\infty}^{t} f_\alpha(s)ds, \quad G_\beta(t) = \int_{-\infty}^{t} g_\beta(s)ds, \quad \text{for } t \in \mathbb{R},$$

and their inverse functions are defined by, with obvious notation y_β,

$$F_\alpha^{-1}(u) = \mathrm{Q}_F(\alpha, x_\alpha, s) \text{ such that } F_\alpha(s) = u, \quad \text{for } u \in (0,1),$$
$$G_\beta^{-1}(v) = \mathrm{Q}_G(\beta, y_\beta, t) \text{ such that } G_\beta(t) = v, \quad \text{for } v \in (0,1).$$

Finally we write $x_F(\alpha, u) = F_\alpha^{-1}(u)$ and $x_G(\beta, v) = G_\beta^{-1}(v)$ the local mass parametrizations on $(0,1)$ of the Q-curves $\mathrm{Q}_F(\alpha)$ and $\mathrm{Q}_G(\beta)$.

The following statement is then straightforward.

Proposition 22. *If* $F \in \bar{\mathcal{F}}$ *then* K_F *has continuous positive density* k_F *on* $(0,1)$, *the quantile curves* $\alpha \to Q_F(\alpha)$ *uniquely determine* F *and the quantile sets* $\alpha \to \mathbb{Q}_F(\alpha)$ *uniquely determine* F.

The generator property on $\mathbb{R}^2$ follows from the fact that the $r.v.$ $x_F(K_F^{-1}(Z_1), Z_2)$ has distribution F. When $F, G \in \bar{\mathcal{F}}$, we again define $\mathcal{G}_F(Z) = x_F(K_F^{-1}(Z_1), Z_2)$ and $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$. The method and results of the previous Section 2 and 3.3 remain valid with slight modifications.

Sections 3.2 and 3.1 also remain true under moment conditions of the kind $\mathbb{E}(A_i(X_i)) < \infty$ and $\mathbb{E}(A_i(Y_i)) < \infty$ with a cost c satisfying $c_i(x - y) \leqslant A_i(x) + A_i(y)$ for $i = 1, 2$ and non negative continuous functions A_1, A_2. This ensures that $\mathbb{E}(c(X, \tau_{FG}(X))) < \infty$.

Up to minor changes this extends to the remainder of distributions in $\overline{\mathbb{F}}$ – intersections of parallel or orthogonal half-spaces.

4 Statistical aspects of the Kendall ordering

4.1 Kendall quantiles, ranks and spacings

The quantiles and ranks built from the generator map $\mathcal{G}_F$ are bivariate statistics that provide a stochastic comparison of distributions within $\overline{\mathbb{F}}$, through the rank square $\mathbb{U}$ and the Kendall geometry.

Definition 23. *For* $z \in \mathbb{U}$, *the* z*-th Kendall quantile point of* $F \in \overline{\mathbb{F}}$ *is*

$$\mathcal{G}_F(z) = x_F(K_F^{-1}(z_1), z_2) = (x_1, x_2) \in Q_F(K_F^{-1}(z_1)).$$

For $x \in \mathcal{R}_X$, *the* x*-th Kendall rank point of* F *is*

$$\mathcal{G}_F^{-1}(x) = \big(K_F(F(x)), F_{F(x)}(t_F(x))\big) = (z_1, z_2) \in \mathbb{U}.$$

The coupling $(\mathcal{G}_F(Z), \mathcal{G}_G(Z))$ affects the same random rank point Z to both marginal $r.v.$'s.

Next we give an alternative limiting expression for $\mathcal{G}_F^{-1}(x)$ that opens access to a natural empirical estimator of z_2. By replacing F with the empirical $d.f.$ we can easily estimate the rank (z_1, z_2), as will be seen in Section 5.

Proposition 24. *The rank point* $z = \mathcal{G}_F^{-1}(x)$ *of* $x \in \mathcal{R}_X$ *satisfies, for* $\mathcal{R}_X^x = \{x' \in \mathcal{R}_X : x'_2 > x_2\}$,

$$z_2 = \lim_{\varepsilon \to 0} \frac{F\left((\mathbb{Q}_F(F(x) + \varepsilon) \backslash \mathbb{Q}_F(F(x))) \cap \mathcal{R}_X^x\right)}{F\left(\mathbb{Q}_F(F(x) + \varepsilon) \backslash \mathbb{Q}_F(F(x))\right)}.$$

Moreover, X *has distribution* F *if, and only if,* $\mathcal{G}_F^{-1}(X)$ *has distribution* U. *For* $z \in \mathbb{U}$ *the* z*-quantile point* $\mathcal{G}_F(z)$ *satisfies (3).*

Proof. From $F\left(\mathbb{Q}_F(F(x)+\varepsilon)\backslash\mathbb{Q}_F(F(x))\right) = K_F(F(x)+\varepsilon) - K_F(x)$ we see by dividing by ε that the denominator tends to $k_F(F(x))$ since $k_F(\alpha) = \partial K_F(\alpha)/\partial\alpha$ by the proof of Proposition 4. Next observe that Definition 9 also applies to the positive measure with distribution $F^x(x') := F(x')$, for $x' \in \mathcal{R}_X^x$. Clearly (8) has solution $\mathrm{Q}_{F^x}(F(x),t) = \mathrm{Q}_F(F(x),t) \in \mathcal{R}_X^x$ for $0 < t < t_F(x) = t_{F^x}(x)$ and $K_{F^x}(F(x)) = F(\mathbb{Q}_F(F(x)) \cap \mathcal{R}_X^x)$ has derivative $k_{F^x}(F(x)) = \int_0^{t_F(x)} f(\mathrm{Q}_F(F(x),t))dt$, according to the arguments of the proof of Proposition 4. As a consequence, the numerator is $K_{F^x}(F(x)+\varepsilon) - K_{F^x}(F(x))$ and the ratio converges to $k_{F^x}(F(x))/k_F(F(x)) = F_{F(x)}(t_F(x))$ which establishes the formula for z_2.

Secondly, if X has $d.f.$ F then $\mathcal{G}_F^{-1}(X)$ has distribution U as $\mathcal{G}_F$ is a one to one map and representation Theorem 8 yields the claimed equivalence. Finally, it holds $\mathbb{P}(\mathcal{G}_F^{-1}(X) \le z) = \mathbb{P}(K_F(F(X)) \le z_1) = z_1$ by (11). By applying Theorem 8 with $X = x_F(K_F^{-1}(Z_1), Z_2)$ we get, for $\alpha_1 = K_F^{-1}(z_1)$,

$$\begin{aligned}
&\mathbb{P}(X_2 \ge \langle \mathcal{G}_F(z), e_2\rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\}\times(0,1)) \\
&= \mathbb{P}(\langle X, e_2\rangle \ge \langle x_F(K_F^{-1}(z_1), z_2), e_2\rangle \mid K_F(F(X)) = z_1) \\
&= \mathbb{P}(\langle x_F(K_F^{-1}(Z_1), Z_2), e_2\rangle \ge \langle x_F(\alpha_1, z_2), e_2\rangle \mid F(X) = \alpha_1) \\
&= \mathbb{P}(\langle x_F(\alpha_1, Z_2), e_2\rangle \ge \langle x_F(\alpha_1, z_2), e_2\rangle \mid Z_1 = z_1) \\
&= \mathbb{P}(Z_2 \le z_2 \mid Z_1 = z_1) \\
&= z_2.
\end{aligned}$$

Therefore (3) holds true. □

The interquantile and spacing notions are also intrinsic to the distribution – with respect to the orthonormal coordinate system. A univariate spacing is the quantile interval associated to an interval of ranks. A bivariate spacing is a curved rectangle depending on F, see Figures 1, 2 for $F = U$. In both cases the spacing represents the generator mass lying between two quantile points and the geometrical shape it takes in the distribution geometry.

Definition 25. *The bivariate spacing of F between ranks $z' \le z''$ is the connex compact set $\{\mathcal{G}_F(z) : z' \le z \le z''\}$.*

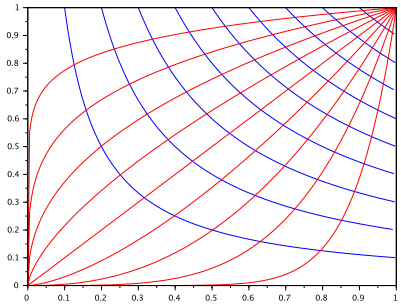

Figure 1: 9 blue α-th Q-curves and 9 red R-curves for $\alpha, z_2 = 0.1 : 0.1 : 0.9$ and $F = U$. These spacings are of different mass (here surface), except those between two consecutive blue curves.

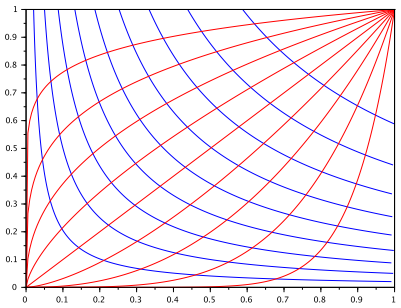

Figure 2: 9 blue z_1-curves and 9 red z_2-curves for $z_1, z_2 = 0.1 : 0.1 : 0.9$ and $F = U$. A partition of 100 spacings of mass 0.01 (here surface).

4.2 One-sided Kendall risk areas

Let assume that the coordinates (X_1, X_2) of X are meaningful statistical quantitative variables. It may happen in applications that having one at least of these quantities too small characterizes a sub-population at risk. We can then propose the following one sided risk aera.

Definition 26. *The Kendall one sided risk area of risk $z_1 \in (0,1)$ is the set $\mathbb{Q}_F(K_F^{-1}(z_1)) = \{x : K_F(F(x)) \le z_1\}$.*

This is the area before the level curve of the distribution F of X at level $\alpha_1 = K_F^{-1}(z_1)$. In other words, each individual of the sub-population at risk z_1 has at most α_1 percent of the population with worse values in both X_1 and X_2. Considering a similar statistical quantitative variable Y with distribution G, we obviously have from Corollary 17 the property that τ_{FG} matches any sub-populations at risk z_1 of X and Y.

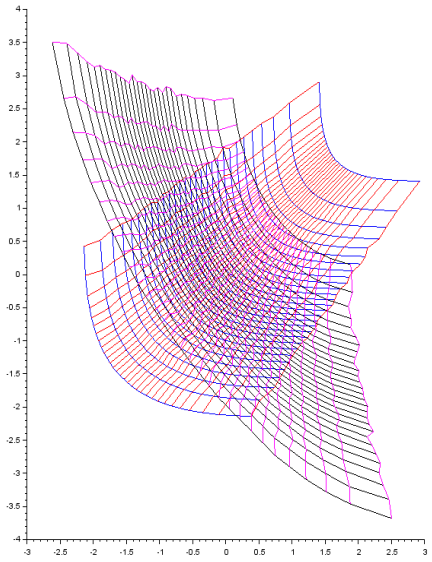

Figure 3: 30×30 (z_1, z_2)-curves estimated with two 300.000 points of bivariate Gaussian $d.f.$ with covariance matrices $\begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$. The ceils have all the same probability, and are organized within the sample ellipsoids.

Example 27. *In a clinical study, similar treatments are administered to two sub-populations with the same risk z_1 in two different human populations, in order to compare responses. Using τ_{FG}, we can partition the two sub-populations in a very similar way - by increasing level of F and G. People in the matched elements of both partitions receive the same dose of medication. The matched partitions use z_2 as well. In Figure 3 we estimate the Kendall areas $\mathbb{Q}_F(K_F^{-1}(k/30))$ and $\mathbb{Q}_G(K_G^{-1}(k/30))$ of two Gaussian populations are the k first bands from below – blue for F and red for G – by using Section 5.*

4.3 Bivariate Kendall tau

Given $X = (X_1, X_2)$ a popular measure of rank correlation between X_1 and X_2 is the univariate [14] Kendall tau

$$\tau(X_1, X_2) = \mathbb{P}\left((X_1 - X_1')(X_2 - X_2') > 0\right) - \mathbb{P}\left((X_1 - X_1')(X_2 - X_2') \leq 0\right)$$

where X and X' are independent with distribution F. If $F \in \overline{\mathbb{F}}$ this reduces to a product of $\{1, -1\}$-valued signs comparison, or equivalently to relative rank comparisons, which is non parametric and robust. Clearly $\tau(X_1, X_2) = 0$ if X_1 and X_2 are independent, and $\tau(X_1, X_2) = 1$ (resp. -1) if, and only if, F is degenerated with $X_2 = F_2^{-1} \circ F_1(X_1)$ (resp. $X_2 = F_2^{-1} \circ (1 - F_1)(X_1)$).

The generator provides a bivariate extension. Let measure the rank correlation of (X, Y) through the generator $r.v.$'s $Z_F = \mathcal{G}_F^{-1}(X)$ and $Z_G = \mathcal{G}_G^{-1}(Y)$. From $\mathcal{G}_F^{-1} = (z_1^F, z_2^F)$ define the signs $s^F(X, X') \in \{1, -1\}^2$,

$$s_j^F(X, X') = 1_{\left\{z_j^F(X) > z_j^F(X')\right\}} - 1_{\left\{z_j^F(X) \leq z_j^F(X')\right\}}, \quad j = 1, 2.$$

Definition 28. *Let $(X, Y) \in \mathbb{R}^2 \times \mathbb{R}^2$ have distribution H with marginals F and G. Given two independent versions (X, Y) and (X', Y') with law H, define the bivariate Kendall correlation to be*

$$k(X, Y) = (k_1(X, Y), k_2(X, Y))$$

where, for $j = 1, 2$,

$$k_j(X, Y) = \mathbb{P}\left(s_j^F(X, X') s_j^G(Y, Y') = 1\right) - \mathbb{P}\left(s_j^F(X, X') s_j^G(Y, Y') = -1\right).$$

If X and Y are independent then $k(X,Y)=(0,0)$. The most extreme correlations $\|k(X,Y)\|_1=2$ are achieved in degenerated cases $Y=\varphi(X)$ that we restrict to $\varphi\in\tau_1(F,G)$. We have $k(X,Y)=(1,1)$ if, and only if, $Y=\tau_{FG}(X)$. Likewise $(1,-1)$, $(-1,1)$ and $(-1,-1)$ are uniquely obtained by $X=\mathcal{G}_F(Z)$ and, respectively, $Y=\mathcal{G}_G(Z_1,1-Z_2)$, $Y=\mathcal{G}_G(1-Z_1,Z_2)$ and $Y=\mathcal{G}_G(1-Z_1,1-Z_2)$.

As illustrated by the following example, $k_j(X,Y)=\tau(z_j^F(X),z_j^G(Y))$ provides a different insight on H than the purely marginal $\tau(X_j,Y_j)$. Indeed, the geometries of F and G are used in each k_j to measure the correlation carried by H only in terms of relative position of (X,Y) in their respective distribution geometry.

Example 29. *Let X_j and Y_j be scores at exam $j=1,2$ – written and oral – in two different lectures. A high $z_1^F(X)$ characterizes a good student in the first lecture, and a high $z_2^F(X)$ indicates a student that performs better at written than at oral in that lecture. Then $k_1(X,Y)$ measures the rank concordance between lectures and $k_2(X,Y)$ between the evaluation type. Separating these two meaningful effects is indeed not possible by using the pairwise marginal Kendall tau $\tau(X_1,Y_1)$, $\tau(X_2,Y_2)$, $\tau(X_1,X_2)$ and $\tau(Y_1,Y_2)$.*

5 Empirical quantile transform maps

5.1 Empirical Kendall distribution

In this section we deal with two i.i.d. samples drawn from two smooth distributions F and G, with size n and m respectively. Write $F_n(x)=\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i\le x}$ the empirical distribution function induced by the sample $(X_1,\dots,X_n)$ from F. Let define the Q-curves, Q-sets and Kendall distribution with respect to F_n mutatis mutandis, simply adding the subscript n : for $0<\alpha<1$,

$$\begin{aligned}\mathrm{Q}_{F,n}(\alpha)&=\left\{x\in\mathbb{R}^2 : F_n(x)=\alpha\right\},\\ \mathbb{Q}_{F,n}(\alpha)&=\left\{x\in\mathbb{R}^2 : F_n(x)\le\alpha\right\},\\ K_{F,n}(\alpha)&=F_n(\mathbb{Q}_{F,n}(\alpha))=\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{F_n(X_i)\le\alpha}.\end{aligned}$$

In the sequel $\alpha=k/n$, $1\le k\le n-1$, so that $K_{F,n}(\alpha)=n^{-1}card\{i:\sum_{j=1}^n \mathbf{1}_{X_j\le X_i}\le k\}$ and, for $F\in\mathbb{F}$, the empirical Q-curves $\mathrm{Q}_{F,n}(k/n)$ are $a.s.$ not empty.

5.2 The empirical Q-curves algorithm

The curves $\mathrm{Q}_{F,n}(\alpha)$ are decreasing step functions with many jumps. However, the algorithm to determine them exactly is straightforward and fast. Assuming $F\in\mathbb{F}$, the first and second coordinates of the sample are $a.s.$ all distinct. One can then sort separately each of the two sets $H_n(\alpha)$ and $V_n(\alpha)$ of the first and second coordinates of the $X_i\in\mathbb{Q}_{F,n}(\alpha)$ – both in increasing order. Denote $F_{i,n}$ the marginal empirical $d.f.$, for $i=1,2$, and $F_{i,n}^{-1}$ their inverse. Use $H_n(\alpha)$ to accelerate the computation of $F_n(X_i)$, $i\le n$, then identify $\mathbb{Q}_{F,n}(\alpha)$. We are ready to draw $\mathrm{Q}_{F,n}(\alpha)$.

Start from the "highest" point of $\mathrm{Q}_{F,n}(\alpha)$ defined to be $(F_{1,n}^{-1}(\alpha),max(V_n(\alpha)))$. Next draw an horizontal line up to the point x having immediately higher first coordinate in $H_n(\alpha)$ then draw from x a vertical line up to the point y having immediately lower second coordinate in $V_n(\alpha)$. Only the upper vertex x is excluded from the stepwise function $\mathrm{Q}_{F,n}(\alpha)$. Indeed x is necessary not an X_i and $x\in\mathrm{Q}_{F,n}(\alpha+1/n)$. Continue the $nK_{F,n}(\alpha)$ steps "$\to x\downarrow y$" until reaching the point $(max(H_n(\alpha)),F_{2,n}^{-1}(\alpha))$ as an $\to x$ or an $\downarrow y$. It is worth to remark that only vertices y can be X_i's and rather few points of $\mathrm{Q}_{F,n}(\alpha)$ belong to the sample – sometimes none. Moreover, the $\mathrm{Q}_{F,n}(\alpha)$ are equal for $k/n\le\alpha<(k+1)/n$.

Example 30. *We shall illustrate a few facts by using samples from mixtures of three or four Gaussian $d.f.$ depicted at Figures 4 and 5. Figures 6, 7, 8 show eight empirical Q-curves, each containing one among eight fixed points on the diagonal $(-4,-3,-2,-1,-0.5,0,1,2)$ for $n=10^3,10^4,10^5$ samples of the three Gaussian mixture. Figure 9 shows the Q-curves crossing the same eight points for the four Gaussian mixture. They clearly differ.*

For α chosen a priori, or given by points, the strips between curves $\mathrm{Q}_{F,n}(\alpha)$ have different empirical probabilities. At the next step we make these probability equal by inverting $K_{F,n}$.

5.3 The empirical quantiles and ranks algorithm

For $z\in\mathbb{U}$ let estimate the quantile point $\mathcal{G}_F(z)=x_F(K_F^{-1}(z_1),z_2)$ by considering the above empirical Q-curve $\mathrm{Q}_{F,n}$ of empirical order $\alpha=K_{F,n}^{-1}(z_1)$. Since the stepwise $\mathrm{Q}_{F,n}(K_{F,n}^{-1}(z_1))$ contains very few sample points – if

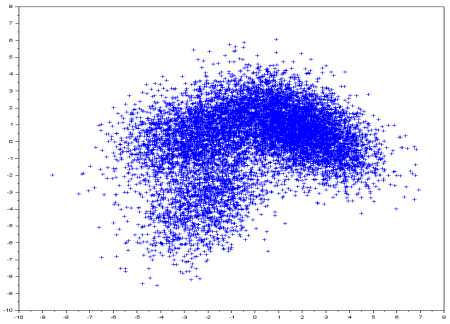

Figure 4: Mixture of 3 Gaussian $d.f.$, subsample 10^4.

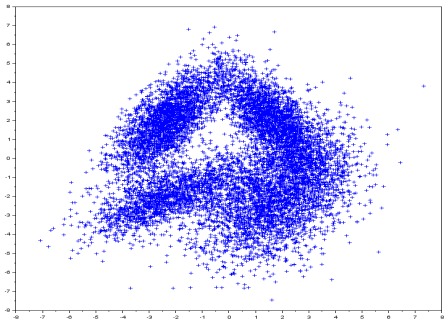

Figure 5: Mixture of 4 Gaussian $d.f.$, subsample 10^4.

any – it is not so obvious how to define an empirical conditional distribution along it that fits F_α along $\mathrm{Q}_F(K_F^{-1}(z_1))$. Unfortunately the R-sets have no property allowing a separate estimation as above for the Q-set $\mathbb{Q}_{F,n}(K_{F,n}^{-1}(z_1))$.

To get an easily computed and mathematically tractable approximation we proceed by enlarging Q-curves to bands between sufficiently separated Q-curves. For this, let select a subset of the Q-curves orders α to define an equipartition of the sample in strips between the selected $\mathrm{Q}_{F,n}(\alpha)$. Hence consider $z_1 = (2i-1)/2p$ for $i = 1, ..., p$ and $z_2 = (2j-1)/2q$ for $j = 1, ..., q$. Taking $p = q \ll n$ yields a regular grid in $\mathbb{U}$ and an equally distributed pavement of the sample. For the asymptotic study of these empirical curves, $p = nh_n$ depends on n and a bandwidth of mass $h_n \to 0$. For sake of simplicity, assume that $n/p \in \mathbb{N}$ and $n/q \in \mathbb{N}$. The number $p \times q$ of sub-strips, and $p + q$ of curves to be estimated, are chosen "reasonably" large in the forthcoming experiments.

Define the quantiles $\alpha_1, \ldots, \alpha_p$ of $K_{F,n}$ verifying $K_{F,n}(\alpha_i) = i/p$. Recall that they can always be taken of the form k/n. A non trivial nonparametric geometrical aspect comes from the fact that α_i are indirect empirical quantiles of $F(X)$ since $F(X_i)$ are not observed. The quantile points we are looking for will be on the "median" curves between $\mathrm{Q}_{F,n}(\alpha_i)$ and $\mathrm{Q}_{F,n}(\alpha_{i+1})$ defined to be $\mathrm{Q}_{F,n}(a_i)$ with $K_{F,n}(a_i) = (2i+1)/p$. Let approximate the conditional quantiles $x_F(K_F^{-1}(z_1), z_2)$ on the curve $\mathrm{Q}_F(\alpha)$ by a random point on $\mathrm{Q}_{F,n}(a_i)$ in the following way.

Denote $S(i)$ the random strip between $\mathrm{Q}_{F,n}(\alpha_i)$ and $\mathrm{Q}_{F,n}(\alpha_{i+1})$ and consider the (outside) virtual point $s(i)$ with coordinates the maxima of the first and second coordinates of the sample points in $S(i)$. An efficient way to order the points in $S(i)$ is to sort them in decreasing order of the angle made by the horizontal axis from $s(i)$ and the lines joining $s(i)$ to the sample points – see Figure 10. By absolute continuity this ordering is $a.s.$ strict. Recalling that all the strips have the same number n/p of sample points, we approximate the quantiles of the conditionnal distribution on $\mathrm{Q}_{F,n}(a_i)$ by intersecting with the z_2-th empirical quantile of the angles. If we are looking for q quantile Q-curve points, we denote them $x(i,j)$ for $z_2 = (2j-1)/q$, $j = 1, \ldots, q$ as on Figure 10. Again for an asymptotic study q should depend on n.

Example 31. *Figure 11 shows 50 empirical Q and R curves for a sample of $\mathcal{N}(0, I_2)$. One can notice that the R-curves of conditionnal, curvilinear quantiles are much less stable than the level Q-curves. This is due to the relatively low number of sample points in each strip and the wellknown difficulty to estimate quantiles even on the real line. Our theoretical study confirms the different rates of approximation of the two types of curves.*

5.4 Quantile transform maps

To obtain an approximation of the quantile transform map τ_{FG} of (4) we compute by the previous algorithm the level curves and conditional quantile curves on two samples of F and G with the same p and q – the sample sizes n and m need not to be equal. If $x(i,j)$ and $y(i,j)$ are the respective conditional quantiles, that we hereafter call the grids, then the approximation $\tau_{n,m}$ of τ_{FG} on the grids is

$$\tau_{n,m}(x(i,j)) = y(i,j). \tag{20}$$

Remind that the grids are obtained as above by intersecting empirical z_1-curves and q z_2-curves. Thus we have built an empirical skeleton of the theoretical quantile transform map by matching these two random grids. One can extend this map to any point x of the convex hull of the F by an interpolation. Alternatively we can use the above algorithms starting from $\alpha = F_n(x)$ then compute the empirical quantile of the angle reaching x – estimated in the angle ordered

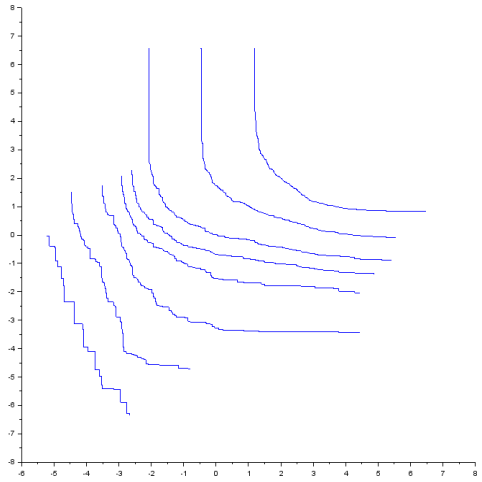

Figure 6: Three Gaussian mixture, $n = 10^3$.

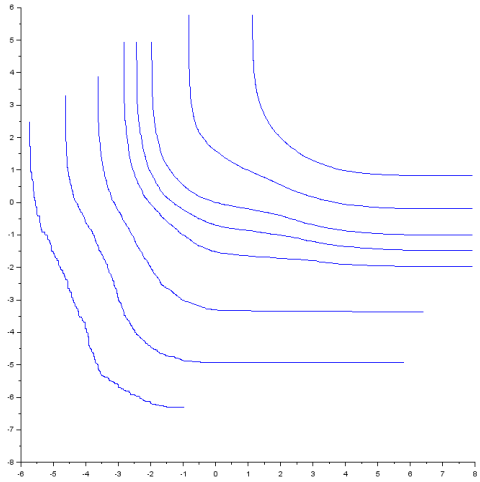

Figure 7: $n = 10^4$.

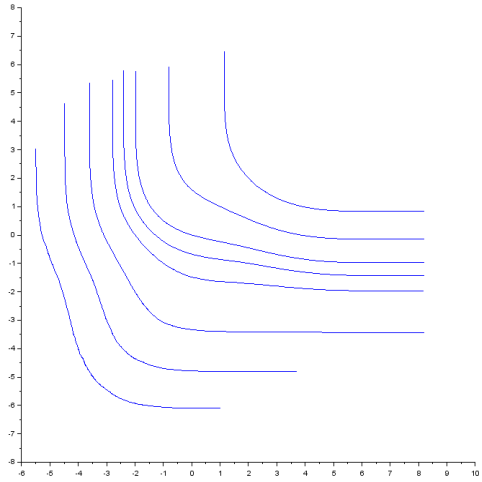

Figure 8: $n = 10^5$.

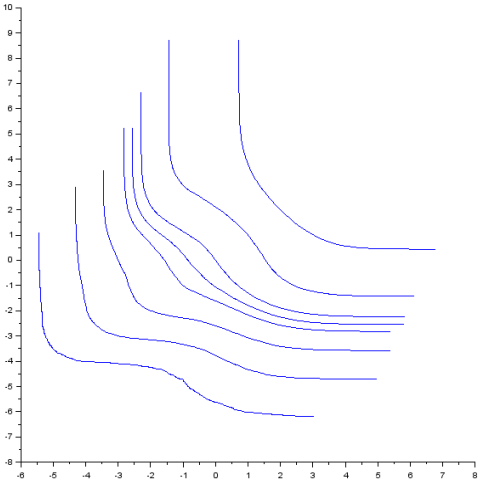

Figure 9: Four Gaussian mixture, $n = 5.10^4$.

strip – and proceed in the same way along the $\mathrm{Q}_{G,n}$ curve of index $K_{G,n}^{-1}(K_{F,n}(\alpha))$, intersected by a line with same estimated angle quantile in the same mass strip. Notice that neither x nor $\tau_{n,m}(x)$ need to be sample points, and $\tau_{n,m}$ can easily be defined everywhere.

5.5 Some numerical examples of quantile transform maps

We restrict ourselves to a few numerical example on $\overline{\mathbb{F}}$, with $n = m = 10^5$. Start with examples in the initial basis b_0.

Example 32. *The goodness-of-fit case $F = G$ allows to "quantify" the numerical error due to our numerical approximating method. The theoretical τ_{FF} is the identity map, and we compute the mean square distance error with the empirical map on the grids. In the previous Gaussian case $\mathcal{N}(0, I_2)$ with two samples of size 10^5 and grids 50×50*

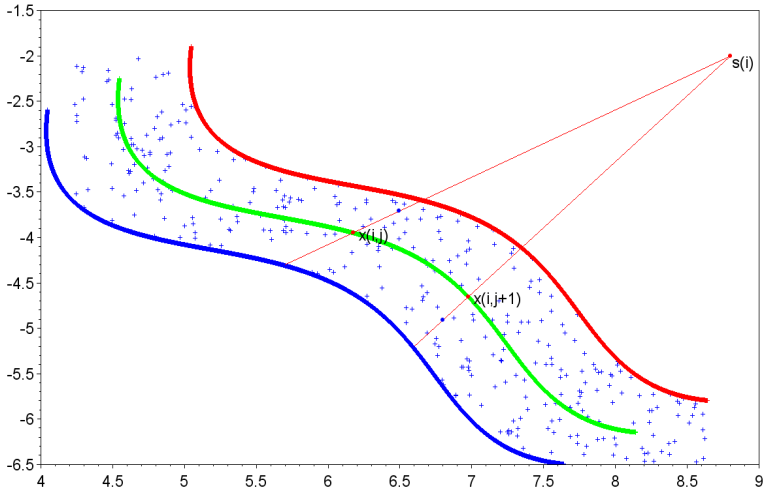

Figure 10: Sample in the strip $S(i)$ between (blue) $\mathrm{Q}_{F,n}(\alpha_i)$ and (red) $\mathrm{Q}_{F,n}(\alpha_{i+1})$, the two (red) points $x(i,j)$, $x(i,j+1)$ on the (green) median curve $\mathrm{Q}_{F,n}(a_i)$ correspond to quantiles of the angles with $s(i)$, ordered top-down, reached by the two (blue) points on the (red) lines.

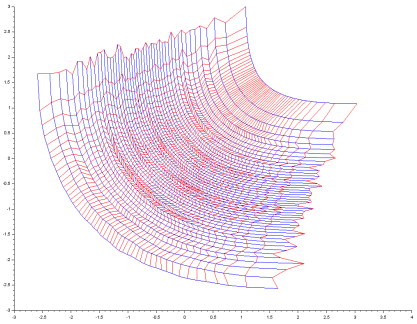

Figure 11: 50 Q-curves (blue) and R-curves (red) of a 10^5 sample of $\mathcal{N}(0, I_2)$.

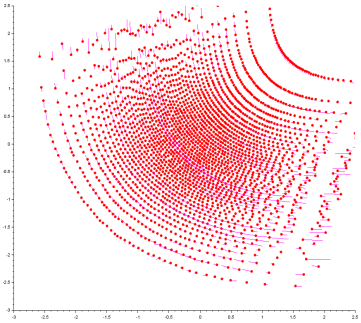

Figure 12: Quantile transform map with grids 50×50, two 10^5 samples of $\mathcal{N}(0, I_2)$: (red) points are the images of the starting grid and the (cyan) lines show the very short move from initial points.

the mean square error is .002 *(the square root is* .045*) which actually is very small – see Figure 12. The fact that the errors follow the curves and seem to have few correlation can be theoretically explained.*

Example 33. *Consider the case when F and G share the same copula. To illustrate Proposition 20 we simulate a* 10^5 *sample of a* $\mathcal{N}(0, \Sigma)$, $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ *and a* 10^5 *sample of a r.v. with the same Gaussian copula and marginals having respective densities* $a|x|e^{-ax^2}$ *with* $a = .05$ *and* $\frac{1}{8}|x|e^{-x/2}$*, that is a symetrized* $\chi^2(4)$*. As expected* $\tau_{n,n}$ *is close to the map estimated on the grid* $x(i,j)$ *by the product of the univariate empirical marginal quantile transforms. Figure 13 shows the pink segments sending* $x(i,j)$ *to the red points* $y(i,j)$*. Figure 14 draws the segments from* $y(i,j)$ *to the image of* $x(i,j)$ *by the product of the marginal quantile transform maps. The mean quadratic error between the two maps is* .027*, that is negligible before the true cost* 17.94 *of the product quantile transform map on the starting grid.*

Example 34. *Another same copula case, with a non classical copula. We simulate two* 10^5 *samples of centered mixtures of Gaussian distributions – with* 4 *and* 3 *components respectively (see Figures 5 and 4). We keep the sample of the*

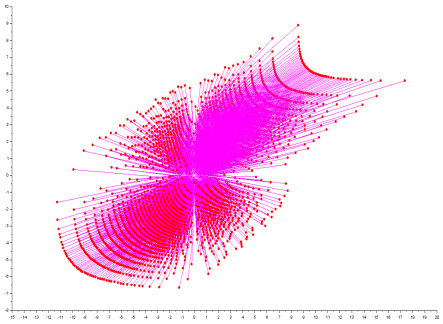

Figure 13: Example with the same Gaussian copula: quantile transform map $x \to \tau_{n,n}(x)$ on the quantile grid 50×50, $n = m = 10^5$.

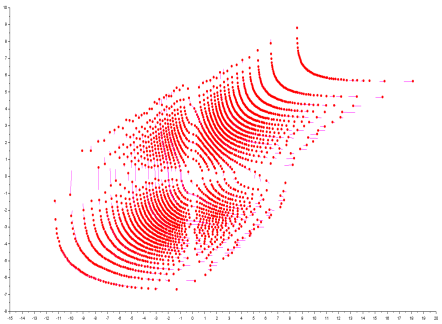

Figure 14: Small errors between images of the initial grid of Figure 13 by $\tau_{n,n}$ – the (red) points – and the product empirical map.

4 Gaussians mixture and draw a second sample of this distribution then transform its marginals onto the empirical marginals of the 3 Gaussians mixtures. We observe that $\tau_{n,n}$ is close to the product of the empirical marginal quantile transform maps up to a mean square error less than .0065, to be compared to the empirical quadratic cost .545.

As a matter of fact, we may use other systems of coordinates.

Example 35. *A polar coordinates case. Let consider two centered radial distributions with the same angle distribution, moreover independent of the radius distribution. To fulfill the positive finite density condition at the center, we simulate a radius r.v. with respective densities $2a\rho e^{-a\rho^2}$ – with $a = .05$ – and $\frac{1}{4}\rho e^{-\rho/2}$ – a $\chi^2(4)$ – and independent angles sharing the density $\frac{1}{4}\sin\theta/2$. Figures 15, 16 show the obtained samples. Figure 17 illustrates that $\tau_{n,n}$ looks radial, that seems confirmed by Figure 18 showing that the images by the radius transform map and the empirical quantile transform map of three thin centered ring almost coincide. The empirical quantile transform map cost is .68 and the radius transform on the grid is .646. By estimating the empirical cost of product of the empirical (polar) marginal quantile transform maps and 50 exact quantiles we get a score 0.643. Hence $\tau_{n,n}$ is efficient despite the fact that it is uninformed of radiality and estimates 50 curves, not numbers.*

6 Proposal of new statistical tools

In order to remove the dependency on the coordinate system given by the marginals, we shall use the rotations r_θ by angle $\theta \in [0, 2\pi)$. Write $F_\theta = F \circ r_{-\theta}$ and $\mathcal{G}_\theta^{-1} = G^{-1} \circ r_{-\theta}$. Minimizing or averaging in θ then makes the proposed tools rotationally equivariant.

6.1 Generator rank

6.1.1 Rank distance, paths and correlation

Any distance d on the rank square $\mathbb{U}$ induces a rank distance between $x, y \in \mathcal{R}_X$ inside the distribution F through

$$r_F(x, y) = d(\mathcal{G}_F^{-1}(x), \mathcal{G}_F^{-1}(y)).$$

If d is the L_2 distance on $\mathbb{U}$, the Q and R curves are treated equally. The rank range $[r_{FF}^-(x, y), r_{FF}^+(x, y)]$ characterizes the proximity of x and y inside F and a rank path can be used to join x to y,

$$t \in [0, 1] \to X_t(x, y) = \int_0^{2\pi} \mathcal{G}_{F_\theta}(t\mathcal{G}_{F_\theta}^{-1}(x) + (1 - t)\mathcal{G}_{F_\theta}^{-1}(y))d_\theta.$$

If x and y share the same R or Q curve, the rank path is a segment along that curve.

Likewise one can define $r_{FG}(x, y) = d(\mathcal{G}_F^{-1}(x), \mathcal{G}_G^{-1}(y))$ if $F, G \in \overline{\mathbb{F}}$, so that $r_{FG}(x, \tau_{FG}(x)) = 0$.

Definition 36. *For F, G in $\overline{\mathbb{F}}$ let the minimal (resp. maximal) rank distance between $x \in \mathcal{R}_X$ and $y \in \mathcal{R}_Y$ be $r_{FG}^-(x, y) = \min_{\theta\in[0,2\pi)} d(\mathcal{G}_{F_\theta}^{-1}(x), \mathcal{G}_{G_\theta}^{-1}(y))$ (resp. $r_{FG}^+(x, y) = \max_{\theta\in[0,2\pi)} d(\mathcal{G}_{F_\theta}^{-1}(x), \mathcal{G}_{G_\theta}^{-1}(y))$) and the mean rank distance be $\overline{r}_{FG}(x, y) = \int_0^{2\pi} d(\mathcal{G}_{F_\theta}^{-1}(x), \mathcal{G}_{G_\theta}^{-1}(y))d\theta$. Clearly $r_{FG}^- \le \overline{r}_{FG} \le r_{FG}^+$ map $\mathbb{R}_X \times \mathbb{R}_Y$ on $\mathbb{R}^+$.*

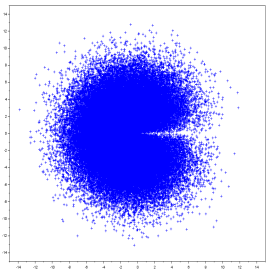

Figure 15: "Pseudo-radial" example, first sample.

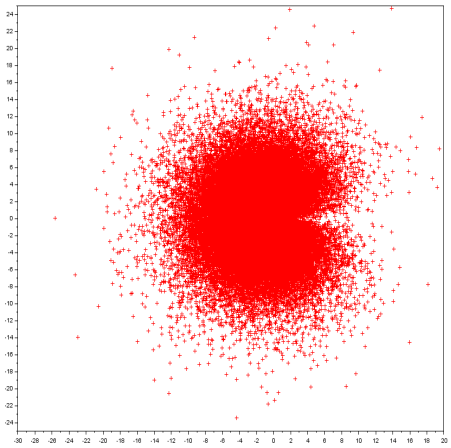

Figure 16: "Pseudo-radial" example, second sample, same scale.

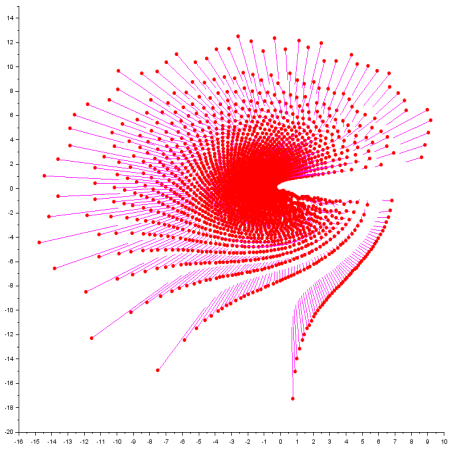

Figure 17: "Radial" example, transportation map τ_{nn} on an initial Kendall quantile grid 50×50, samples sizes 10^5.

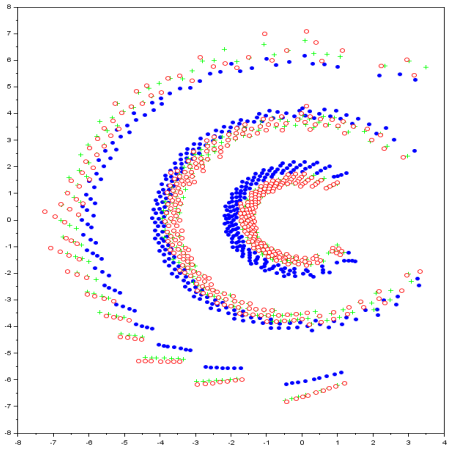

Figure 18: (blue) $\bullet$ are sent to (red) o by $\tau_{n,n}$ map and to (green) $+$ by radius transform map.

Hence $\overline{r}_{FG}(x,x)$ quantifies the different positioning of x in the planar distributions F and G. The statistically meaningful situation is when (F,G) are marginals of a distibution in $\mathbb{R}^4$. An alternative to the natural bivariate Kendall rank correlation of Definition 28 could be based on the following quantities.

Definition 37. *Let $(X,Y) \in \mathbb{R}^2 \times \mathbb{R}^2$ have distribution H with marginals F and G. Rank correlation coefficients associated to the rank distance d on $\mathbb{U}$ are*

$$R_d(X,Y) = \mathbb{E}(r_{FG}(X,Y))) = \mathbb{E}(d(\mathcal{G}_F^{-1}(X), \mathcal{G}_G^{-1}(Y))),$$
$$\overline{R}_d(X,Y) = \mathbb{E}(\overline{r}_{FG}(X,Y)) = \int_0^{2\pi} R_d(r_\theta(X), r_\theta(Y))d\theta.$$

6.1.2 Partitions, modes and clustering

The above rank distances refer to the amount of probability mass lying between to distant points or aeras, using the generator geometry. In the same vein, the point $\mathcal{G}_F(1/2, 1/2)$ may be called the median point in the current coordinate system, as the intersection of the unique Q and R curves separating the support in two equiprobable subsets.

More generally, a nice feature of $\mathcal{G}_F$ is to provide tesselations of $\mathcal{R}_X$ in ceils with desired probability. Start from any partition A_i of the unit square $\mathbb{U}$, with surfaces $|A_i|$. For instance in squares, triangles or hexagons with same surface, or a mixed configuration. The quantile ceils $\mathcal{G}_F^{-1}(A_i)$ have probabilities $|A_i|$ and sometimes surprising characteristic shapes adapted to the Kendall geometry characterizing F – see Figure 19. This can be used to visually compare how two distributions differ in terms of generator quantiles – in more difficult situations than Figure 20.

This could be usefull in descriptive data analysis, by choosing the principal component basis in which the partition minimizes a shape criterium – for instance closer to the partition of a Gaussian sample of the same size.

Example 38. *The empirical partition $\mathcal{G}_{F,n}(A_i)$ could be used to build balanced random forests, in a fast and data-driven way. For this the random refinement of the forest could be done on $\mathbb{U}$ directly by horizontal or vertical separations, cutting $\mathcal{G}_{F,n}(A_i)$ in two parts according to either an empirical Q-curve or and empirical R-curve.*

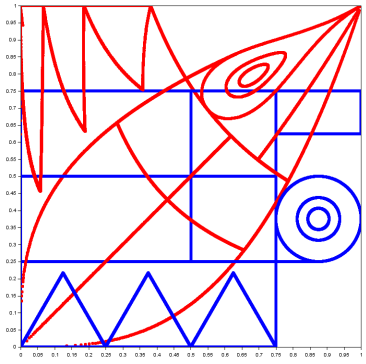

Figure 19: The curved (red) partition is the $\mathcal{G}_U^{-1}$ image of the line-circle (blue) partition of ranks in the Kendall geometry of the uniform distribution U as computed at Example 12. The surface of ceils is preserved.

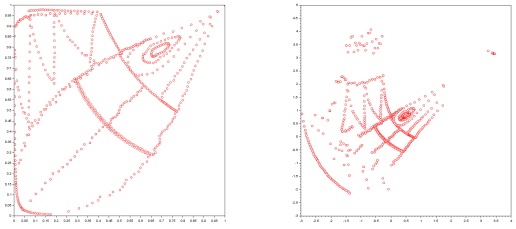

Figure 20: Empirical uniform and gaussian quantile curves, for $n = 2.10^5$ samples, for some points of the blue ranks of Figure 19. The unbounded support disrupts the extremes – ranks z_1 or z_2 close to 0 or 1.

In clustering, meaningful points are modes. A way to estimate modal areas is to locate the minimum surface $|A_{i,j}|$ among a partition of spacings $A_{i,j}$ of same probability, typically the spacings delimited by k Q-curves and k R-curves. The empirical versions again need to relate k to the sample size n.

Definition 39. *Assume $F \in \overline{\mathbb{F}}$ and f has a unique maxima. The order k generator mode of F is $M_k = \mathcal{G}_F(i'/k, j'/k)$ where $(i', j') = \text{Argmin}\,|A_{i,j}|$ and $A_{i,j} = \{\mathcal{G}_F(z_1, z_2) : [kz_1] = i, [kz_2] = j\}$.*

Clearly, M_k tends to the unique mode as $k \to \infty$, whatever the basis choice for marginals. This could easily be extended to a multimodal distribution by using local minima of the surface of ceils. If multiple modes are detected through the smallest distribution spacings then rank paths may link them to picture the probabilistic geometry of F.

6.2 Generator depth

6.2.1 Global depth sets through quantile shapes

Many data analysis methods aim to draw contours or increasing sets to picture some inward-outward ordering of multivariate distributions. They can be based on depth, directional quantiles, regression quantile, or projections. More recent approaches use the implicit optimal transport for a given cost, such as [10] and [12]. We propose to use any increasing sequence of generator spacings to define inward-outward quantile sets with chosen probabilities and shapes depending on the distribution geometry.

Definition 40. *Let $\mathcal{U} = \{U_\alpha : \alpha \in (0,1)\} \subset \mathbb{U}$ be an increasing collection of rank sets such that $\alpha = |U_\alpha|$. The associated quantile sets are $\mathcal{G}_F(\mathcal{U})$. The minimal (resp. maximal) quantile shapes are $\mathcal{G}_F^-(U_\alpha) = \bigcap_0^{2\pi} \mathcal{G}_{F_\theta}(U_\alpha)$ (resp. $\mathcal{G}_F^-(U_\alpha) = \bigcup_0^{2\pi} \mathcal{G}_{F_\theta}^-(U_\alpha)$), for $\alpha \in (0,1)$.*

Clearly $\mathcal{G}_F^-(U_\alpha) \subset \mathcal{G}_F(U_\alpha) \subset \mathcal{G}_F^+(U_\alpha)$ hence $P(\mathcal{G}_F^-(U_\alpha)) \leq P(\mathcal{G}_F(U_\alpha)) = \alpha \leq P(\mathcal{G}_F^+(U_\alpha))$ for any $\alpha \in (0,1)$.

The arbitrary central area or point corresponds to α close to 0, a trimmed area corresponds to α close to 1.

As an illustration in the case of an increasing collection $\mathcal{U}$ of circles centered on $\mathcal{G}_F(1/2, 1/2)$ for a 2.10^5 sample of uniform distribution, Figure 21 shows the empirical contours and Figure 22 compares empirical contour points with exact ones.

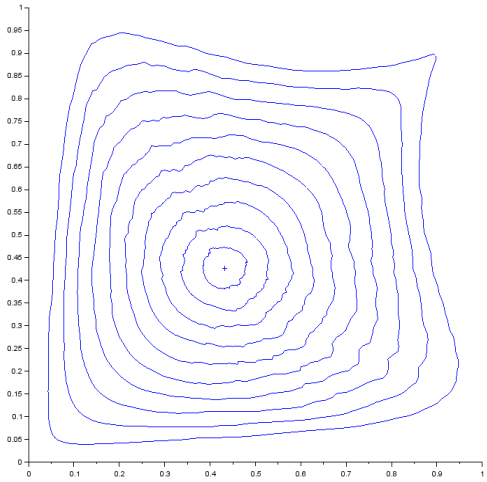

Figure 21: Empirical uniform inward-outwards quantile shapes, $n = 2.10^5$.

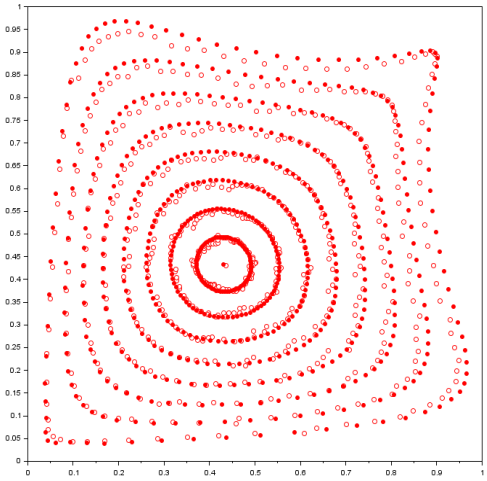

Figure 22: Empirical (circles) and true (red) countour points.

6.2.2 Local depth and contours

A depth value aims to quantify whether a point x is close to the main mass concentrations or not. Depth level sets induce contours that can be interpreted. Various depth notions have been studied and used in data analysis – see [17], [25], [21], [22], [24], [11], among many others. Tukey or simplicial depth, as well as variants such as quantile surfaces [AB], are connected to a function attached to each point – see [20]. As the bidimensional contours or depth levels can be drawn, after some projection, planar depth notions have a strong visual practical interest. Let define local depth and contour notions based on the explicit Kendall geometry of F driven by the generator.

Let define the $\mathcal{G}$-depth relatively to a central point $y \in \mathcal{R}_X$, typically $\mathcal{G}_F(1/2, 1/2)$ or a symetry point, to be $1/d(\mathcal{G}_F^{-1}(x), \mathcal{G}_F^{-1}(y))$. This puts uniformly bounded small values at boundaries and infinite value at y – take the inverse ratio to revert the scoring. This definition easily extends to a depth value relatively to a set $\mathcal{Y}$. Since F is a measure in the euclidean plane we give a basis free definition of relative local depth.

Definition 41. *Let $F \in \overline{\mathbb{F}}$. The $\mathcal{G}$-depth relative to $y \in \mathcal{R}_X$ is*

$$x \in \mathcal{R}_X \to D_F(x|y) = \int_0^{2\pi} \frac{1}{d(\mathcal{G}_{F_\theta}^{-1}(x), \mathcal{G}_{F_\theta}^{-1}(y))} d\theta.$$

The $\mathcal{G}$-local depth relative to $\mathcal{Y} \subset \mathcal{R}_X$ is

$$x \in \mathcal{R}_X \to D_F(x|\mathcal{Y}) = \inf_{y \in \mathcal{Y}} D_F(x|y).$$

If $card(\mathcal{Y}) < \infty$ the $\mathcal{G}$-attractor function of $\mathcal{Y}$ is

$$x \in \mathcal{R}_X \to A_F(x|\mathcal{Y}) = \operatorname*{Argmin}_{y \in \mathcal{Y}} D_F(x|y).$$

In a descriptive statistics perspective, $\mathcal{Y}$ are mass concentration centroïds, such as modes of a multimodal density f or sub-population reference points in a mixture of sub-populations. These centers can be estimated as in 6.1.2.

The $\mathcal{G}$-depth sets $\{x : D_F(x|y) \geq \gamma\}$ and $\mathcal{D}_\gamma(\mathcal{Y}) = \{x : D_F(x|\mathcal{Y}) \geq \gamma\}$ are strictly increasing for inclusion as γ increases. We can deduce from Definition 41 confident regions around specific points $\mathcal{Y}$ to be central in the region.

The $\mathcal{G}$-contours $\mathcal{C}_\gamma(\mathcal{Y}) = \{x : D_F(x|\mathcal{Y}) = \gamma\}$ are closed, continuous, nested curves describing the mass distribution in a similar way as density level sets. They fit what is expected from probability contours – see Figure 23.

Local depths and contours can again serve as refined trimming – even inside the data for very small density areas.

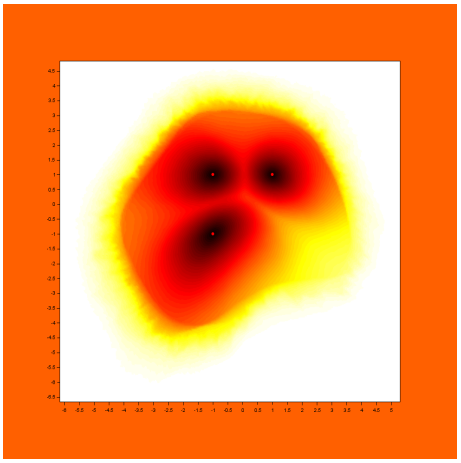

Figure 23: Depth contours obtained for the Gaussian mixture of Figure 4 by using 128 angles and the empirical version of Definition 41.

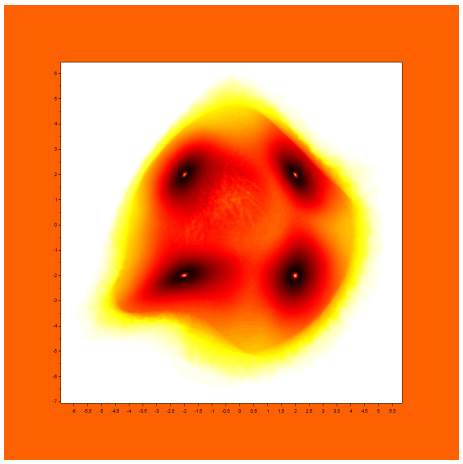

Figure 24: Empirical depth contours for the mixture of Figure 5 by using 64 angles.

6.3 Contrasts

6.3.1 Kendall geometry contrasts

The contrast R_d and $\overline{R}_d$ of Definition 37 are evaluated in the rank geometry to compare correlated $r.v.$'s. Here we compare distributions themselves, in the quantile geometry.

Definition 42. *Given a coordinatewise cost c as in Corollary 17, consider the contrast between F and G in $\mathbb{F}$,*

$$C(F,G) = \int_{\mathbb{U}} c(\mathcal{G}_F(z), \mathcal{G}_G(z))dz = \mathbb{E}_X(c(X, \tau_{FG}(X))).$$

Note that $C(F,G) = 0$ if, and only if, $F = G$. The contrast C may be statistically meaningful by choosing c_1 and c_2 according to the marginals (X_1, X_2) and (Y_1, Y_2) in the case of distributions F and G of similar type.

If one compare two probability measures without focusing on marginals, there is no natural basis and it makes sense to consider all of them. It is then natural to choose a symetric cost function c and make use of

$$C^-(F,G) = \min_{\theta\in[0,2\pi)} C(F_\theta, G_\theta), \quad C^+(F,G) = \max_{\theta\in[0,2\pi)} C(F_\theta, G_\theta), \quad \overline{C} = \int_{(0,2\pi)} C(F_\theta, G_\theta)d\theta.$$

If one have initial coordinates and a c making statistical sense, one can use

$$C^*(F,G) = \int_{(0,2\pi)} \int_{\mathbb{U}} c(r_\theta^{-1}(\mathcal{G}_{F_\theta}(z)), r_\theta^{-1}(\mathcal{G}_{G_\theta}(z)))dzd\theta$$

that always computes in the meaningful initial basis the transport cost of the quantile transform in each rotated basis.

Observe that the $r.v.$ $Y_\theta = r_\theta^{-1}(\tau_{F_\theta G_\theta}(r_\theta(X)))$ has distribution G for all θ. If c is the quadratic cost, the choice of the orthogonal basis is unimportant. In a goodness-of-fit test context, a robust decision could then be based on

$$C_2^*(F,G) = \min_{\theta\in[0,2\pi)} \int_{\mathbb{U}} ||\mathcal{G}_{F_\theta}(z) - \mathcal{G}_{G_\theta}(z)||_2^2 dz = \min_{\theta\in[0,2\pi)} \mathbb{E}(||X_\theta - Y_\theta||_2^2). \tag{21}$$

6.3.2 Empirical contrasts

In a sample comparison setting, substitute $\mathcal{G}_{F,n}$ to $\mathcal{G}_F$ in the above contrasts. In particular,

$$C_{n,m}(F,G) = \int_{\mathbb{U}} c(\mathcal{G}_{F,n}(z), \mathcal{G}_{G,m}(z))dz \quad \text{or} \quad C^*_{2,n,m}(F,G) = \min_{\theta\in[0,2\pi)} \int_{\mathbb{U}} ||\mathcal{G}_{F_\theta,n}(z) - \mathcal{G}_{G_{\theta,m}}(z)||_2^2 dz. \tag{22}$$

Likewise, in order to estimate $C(F,G)$ for F unknown and a targeted G known, simply replace $\mathcal{G}_{G,m}$ with $\mathcal{G}_G$ in (22) – or estimated with a very large m-sample. In the algorithms used for our illustrations, the integrals are Riemann sums on a regular grid z_k of $\mathbb{U}$.

The empirical contrasts $C_{n,m}$ and $C^*_{2,n,m}$ are robust since changing drastically a few sample points won't change very much the empirical bivariate quantile geometry explicited at Section 5. We also insist that $C_{n,m}(F,G) \neq C(F_n, G_m)$ since C is only defined on $\mathbb{F}\times\mathbb{F}$, and for $C_{n,m}$ the probabilistic and geometric construction of empirical Q and R curves is a smoothed version of those of F_n and G_m using parameters $p = q \ll n$.

Moreover, $C_{n,m}$ relies on the non parametric estimation of the quantile geometry itself, which extracts a quantile-rank type information from the evaluation the $d.f.$ indexed by quadrants. As clearly shown in [3] the empirical Q-curves are more fastly learned than the R-curves due to the local, curvilinear information the latter convey. However, the CLT rate for $C_{n,m}(F,G)$ is still the same as for other global scores based on quadrants, such as $\sup|F_n - G_m|$, and the closed form limiting variance explicitly depends on the geometries of F and G.

6.3.3 Classification

The Q-curves and R-curves can be used in classification of bivariate distributions. Assume one wants to classify a collection of rather large bivariate samples, each produced independently – like a series of images or multiple measurements. The contrasts of 6.3.1 and 6.3.2 are efficient, especially if c and the marginals are meaningful enough to distinguish these bivariate distributions through the Kendall ordering. This is the case for instance in genetics or structural biology data sets based on two angles in the dna of disordered proteins – see e.g. [9].

The j-th sample of size $n(j)$ is then assumed to have an unknown $d.f.$ F^j that we can mutually compare and classify through the contrast matrix $C(F^j_{n(j)}, F^{j'}_{n(j')})$ of Definition 42. In order to avoid computing the full matrix, an approach could be to first classify the functional data type Kendall $d.f.$'s $K_{F^j,n(j)}$ that are quickly estimated. This first step could simply be based on $\sup_{(0,1)}|K_{F^j,n(j)} - K_{F^{j'},n(j')}|$. Then, for similar empirical Kendall $d.f.$'s the contrast acts in a second step as a refinement measure to further sub-classify. Indeed, $C_{n,m}$ of (22) really takes into account the actual geometry of the Q-curves and R-curves, that are possibly very different even if $K_{F^j} = K_{F^{j'}}$ at the first step.

Alternatively the first step could already take into account the differences between the Q-curves of F^j and $F^{j'}$. For instance by comparing the surfaces of their respective bands between fixed areas at risk $0 < z^1_1 < ... < z^k_1 < 1$ in the sense of Definition 26. In a second step again sub-classify according to the contrast C and find cluster centers in a k-means way, thus completing the information through the R-curves.

Such a classification could be made faster by comparing quantiles $\mathcal{G}_F(z)$ only for a small finite subset of ranks $z \in \mathcal{Z}_0 \subset \mathbb{U}$. Choosing a few crucial ranks $\mathcal{Z}_0$ to match, the fastly computed contrast becomes

$$C_{0,n,m}(F,G) = \frac{1}{card(\mathcal{Z}_0)} \sum_{z\in\mathcal{Z}_0} c(\mathcal{G}_{F,n}(z), \mathcal{G}_{G,m}(z)). \tag{23}$$

6.4 Tests

6.4.1 Goodness of fit tests

The above contrasts naturally induce goodness of fit test statistics. To test if $F = F_0$ then use $C_{0,n,m}(F,F_0)$ of (23) for a very large m or the stronger

$$\Gamma_n = \frac{1}{k}\sum_{\theta\in\Theta_k} \frac{1}{card(\mathcal{Z}_0)} \sum_{z\in\mathcal{Z}_0} c(r_\theta \circ \mathcal{G}_{F_\theta,n}(z), r_\theta \circ \mathcal{G}_{(F_0)_\theta}(z)).$$

where $\Theta_k \subset [0,2\pi)$ is a finite collection of k angles and $\mathcal{G}_{F_\theta,n}$ is the empirical generator algorithm applied to the rotated sample $r_{-\theta}(X_i)$. At the first use of the test, $G_{(F_0)_\theta}$ is not explicit and should be computed from a huge sample, only for $(\theta,z) \in \Theta_k \times \mathcal{Z}_0$. To test if $F \in \mathcal{F}_0$ one can first almost minimize $\|K_{F,n} - K_{F_0}\|$ over $\mathcal{F}_0$ – assuming the K_{F_0} estimated, classified or known – then minimize $C_{0,n,m}(F,F_0)$ among the almost minimizers F_0.

6.4.2 Bivariate copula comparaison test

Copula is an oriented notion, rotations are not needed. By Proposition 20, comparing the quantile transform map and the marginal quantile transport maps allows to detect when two bivariate distributions share the same copula.

Write F_1 and F_2 the marginals $d.f.$ of F. Consider a cost c as in Corollary 17 and the finite grid $\mathcal{Z}_{p,q} \subset \mathbf{U}$ of ranks used in the generator algorithm – typically uniform if no prior information is available. Hence $card(\mathcal{Z}_{p,q}) = pq$. Let $\widehat{\tau}_{n,m} = (G^{-1}_{1,m} \circ F_{1,n}, G^{-1}_{1,m} \circ F_{2,n})$ be the empirical marginal quantile transforms estimating $(\tau_{F_1G_1}, \tau_{F_2G_2})$. Write $\mathcal{G}_{\widehat{G},n,m}(z)$ the empirical generator computed by using the sample $\widehat{Y}_i = \widehat{\tau}_{n,m}(Y_i)$, $i = 1, ..., m$. Define

$$\Gamma_{n,m} = \frac{1}{pq} \sum_{z \in \mathcal{Z}_{p,q}} c(\mathcal{G}_{F,n}(z) - \mathcal{G}_{\widehat{G},n,m}(z)). \tag{24}$$

We compute $\Gamma_{n,m}$ at Example 34 for the quadratic cost. By Proposition 20 the theoretical version of $\Gamma_{n,m}$ is 0 if F and G have same copula. The proposed copula test consists in accepting the same copula hypothesis whenever $\Gamma_{n,m}$ is small enough. Our theoretical results in [3] can be applied to derive a CLT with explicit limit variance for $\sqrt{n}\Gamma_{n,m}$ by choosing (n, m, p, q) and $\mathcal{Z}_{p,q}$ properly as $n \to \infty$ under the null hypothesis, and to determine the exact limiting behaviour under the different copula alternative.

6.4.3 Distribution anomaly multi-tests

In the same spirit as 6.3.3, given q bivariate typical distributions $G_1, ..., G_q$ one may need to decide whether a new bivariate sample is typical or outlier. Evaluate the $\mathcal{G}_{G_k}$ by using large enough reference samples providing empirical $d.f.$'s $G_{m_1}, ..., G_{m_q}$. Then locate the new sample by comparison based on a contrast from Section 6.3.1. For instance, if the marginal make sense in the statistical problem, use the empirical quantities $C_{n,m}(F, G_1), ..., C_{n,m}(F, G_q)$ using only the reference samples and the new one, with rather few ranks (p, q) in the empirical quantile transform algorithm of Section 5. To sharpen the probability measure comparision use $C^*_{2,n,m}$ instead, for a longer computation time.

The proposed multitest rejects the typical distribution hypothesis if the contrasts are jointly significantly large. To make this statement precise, [3] could be applied to establish a joint CLT for such complicated events. Sharp p-values can then be evaluated for detecting anomalies.

6.5 Concluding comment

In this paper we define new bivariate quantiles and ranks in the spirit of their univariate definition, using the $d.f$ F of X as unique tool. One reason for this choice is that the induced geometry represented by two families of curves is uniquely and meaningfully related to the probability distribution. Another reason is the purely non parametric nature, fast computation speed and low complexity of the algorithm – having only two parameters (p, q). The third reason is that we can take advantage of the Gaussian strong approximation of F by its empirical counterpart F_n to establish the exact asymptotic behaviour of the empirical quantiles and ranks, entirely determined by the geometry – see [3]. Therefore the bivariate rank to quantile transform maps $\tau_{n,m}$ and τ_{FG} open access to new methods in data analysis (depth, contours, classification), statistical inference (contrasts, rank correlation, modes, trimming) and tests (copula, comparison, outliars), as suggested in Section 6.

Considering the coordinate system provided by the marginal $r.v.$'s is somehow a natural way of thinking in statistics, however from a probabilistic point of view the dependency on this basis is a limitation. Actually our generator of X consists in conditioning X by the $r.v.$ $F(X)$ and considering the couple $\big(F(X), X|F(X)\big)$. This is a special case of the wellknown Rosenblatt method, with $F(X)$ as the first $r.v.$ Alternative conditionings should be worked out to overcome the basis dependency. Quantiles and a quantile transform map similar to τ_{FG} could be derived from the couple $(\Phi(X), X|\Phi(X))$ with a real valued function Φ whose level curves characterize F and are independent of the basis chosen for computations. A restriction on the possible Φ is to be easily estimated from a n-sample with the usual rate $\sqrt{n}$.

References

[1] A. Ahidar-Coutrix and P. Berthet. Convergence of Multivariate Quantile Surfaces. arXiv:1607.02604v2, 2016.

[2] P. Barbe, C. Genest, K. Goudhi, and B. Rémillard. On kendall's process. J. Multivariate Anal., 58, 1996.

[3] P. Berthet and J-C. Fort. Convergence of bivariate quantile, rank and transport empirical processes. arXiv:, 2023.

[4] P. Chaudhuri. On a geometric notion of quantiles for multivariate data. Journal of the American Statistical Association, 91(434):862–872, 1996.

[5] V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge–Kantorovich depth, quantiles, ranks and signs. The Annals of Statistics, 45(1):223 – 256, 2017.

[6] J.A. Cuesta Albertos, L. Ruschendorf, and A. Tuero-Diaz. Optimal coupling of multivariate distributions and stochastic processes. J. Multivariate Anal., 46, 1993.

[7] J.H.J Einmahl and D.M. Mason. Generalized Quantile Processes. The Annals of Statistics, 20(2):1062 – 1078, 1992.

[8] C. Genest and L.-P. Rivest. Statistical inference procedures for bivariate archimedean copulas. J.A.S.A., 88, 1993.

[9] Javier González-Delgado, Alberto González-Sanz, Juan Cortés, and Pierre Neuvial. Two-sample goodness-of-fit tests on the flat torus based on wasserstein distance and their relevance to structural biology. Electronic Journal of Statistics, 17(1), jan 2023.

[10] M. Hallin. Measure transportation and statistical decision theory. Annual Review of Statistics and Its Application, 2021.

[11] M. Hallin, Z. Lu, D. Paindaveine, and M. Šiman. Local bilinear multiple-output quantile/depth regression. Bernoulli, 21(3), 2015.

[12] M. Hallin and G. Mordant. Center-outward multiple-output lorenz curves and gini indices a measure transportation approach. arXiv, 2022.

[13] M. Hallin, D. Paindaveine, and M. Šiman. Multivariate quantiles and multiple-output regression quantiles: From L1 optimization to halfspace depth. The Annals of Statistics, 38(2):635 – 669, 2010.

[14] M.G. Kendall. A new measure of rank correlation. Biometrika, 1938.

[15] V. I. Koltchinskii. M-estimation, convexity and quantiles. The Annals of Statistics, 25(2):435 – 477, 1997.

[16] L. Kong and I. Mizera. Quantile tomography: Using quantiles with multivariate data. Statistica Sinica, 22(4):1589–1610, 2012.

[17] R. Liu. On a notion of data depth based on random simplices. Annals of Statistics, 18:405–414, 1990.

[18] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman. Plugin estimation of smooth optimal transport maps. arXiv, 2022.

[19] T. Manole and J. Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. arXiv, 2021.

[20] I. Mizera. On depth and deep points: a calculus. The Annals of Statistics, 30(6):1681 – 1736, 2002.

[21] K. Mosler. In Robustness and Complex Data Structures, chapter 12 Depth statistics. Springer Berlin, 2014.

[22] K. Mosler and P. Mozharovskyi. Choosing Among Notions of Multivariate Depth Statistics. Statistical Science, 37(3):348 – 368, 2022.

[23] R. B. Nelsen, J. J. Quesada-Molina, J. A. Rodriguez-Lallena, and M. Ubeda Flores. Kendall distribution functions. Statist. Probab. Lett., 65(3), 2003.

[24] Y. Zuo. Multidimensional medians and uniqueness. Computational Statistics & Data Analysis, 66:82–88, 2013.

[25] Y. Zuo and R. Serfling. General notions of statistical depth function. Annals of Statistics, 28:461–482, 2000.