On Distribution and Quantile Functions, Ranks and Signs in R^d

Marc Hallin
SBS-EM, ECARES, Université libre de Bruxelles

September 2017

ECARES working paper 2017-34

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

On Distribution and Quantile Functions, Ranks and Signs in $\mathbb{R}^d$

a measure transportation approach

Marc Hallin *
ECARES, Université libre de Bruxelles

Abstract

Unlike the real line, the d-dimensional space $\mathbb{R}^d$, for $d \geq 2$, is not canonically ordered. As a consequence, such fundamental and strongly order-related univariate concepts as quantile and distribution functions, and their empirical counterparts, involving ranks and signs, do not canonically extend to the multivariate context. Palliating that lack of a canonical ordering has remained an open problem for more than half a century, and has generated an abundant literature, motivating, among others, the development of statistical depth and copula-based methods. We show here that, unlike the many definitions that have been proposed in the literature, the measure transportation-based ones introduced in Chernozhukov et al. (2017) enjoy all the properties (distribution-freeness and the maximal invariance property that entails preservation of semiparametric efficiency) that make univariate quantiles and ranks successful tools for semiparametric statistical inference. We therefore propose a new *center-outward* definition of multivariate distribution and quantile functions, along with their empirical counterparts, for which we establish a Glivenko-Cantelli result. Our approach, based on results by McCann (1995), is geometric rather than analytical and, contrary to the Monge-Kantorovich one in Chernozhukov et al. (2017) (which assumes compact supports, hence finite moments of all orders), does not require any moment assumptions. The resulting ranks and signs are shown to be strictly distribution-free, and maximal invariant under the action of a class of (order-preserving) transformations generating the family of absolutely continuous distributions; that maximal invariance, in view of a general result by Hallin and Werker (2003), is the theoretical foundation of the semiparametric efficiency preservation property of ranks. The corresponding quantiles are equivariant under the same transformations.

AMS 1980 subject classification: 62M15, 62G35.
Keywords: Multivariate distribution function; multivariate quantiles, multivariate ranks; multivariate signs; multivariate order-preserving transformation; Glivenko-Cantelli; invariance/equivariance; gradient of convex function.

*The original version of this manuscript was completed while the author was visiting the University Carlos III in Madrid thanks to a Cátedra de Excelencia generously funded by the Banco Santander. A thorough revision was prepared during a stay at the Isaac Newton Institute for Mathematical Sciences in Cambridge within its 2018 programme on Statistical Scalability [EPSRC grant no EP/K032208/1]. The warm hospitality of the Department of Statistics in Madrid-Getafe and the Statistical Laboratory in Cambridge, as well as the support of the Banco Santander and Isaac Newton Institute, are gratefully acknowledged.

1 Introduction

Unlike the real line, the real space $\mathbb{R}^d$, for $d \geq 2$, is not canonically ordered. As a consequence, such fundamental concepts as quantile and distribution functions, which are strongly related to the ordering of the observation space, and their empirical counterparts—ranks and empirical quantiles—playing, in dimension $d = 1$, a fundamental role in statistical inference, do not canonically extend to dimension $d \geq 2$.

Of course, a classical concept of distribution function—the familiar one, based on marginal orderings—does exist. That concept, from a probabilistic point of view, does the job in the sense of characterizing the underlying distribution. However, the corresponding quantile function does not mean much (see, e.g., Genest and Rivest (2001)), and the corresponding empirical versions (related to their population counterparts via a Glivenko-Cantelli result) do not possess any of the properties that make them successful inferential tools in dimension $d = 1$.

That observation about traditional multivariate distribution functions is not new: palliating the lack of a "natural" ordering of $\mathbb{R}^d$—hence, defining statistically sound concepts of distribution and quantile functions—has remained an open problem for more than half a century, and has generated an abundant literature that includes, among others, the theory of copulas and the theory of statistical depth.

A number of most ingenious solutions have been proposed, each of them extending some chosen features of the well-understood univariate concepts, with which they coincide for $d = 1$. Coinciding, for $d = 1$, with the classical concepts obviously is important, but it is hardly sufficient for qualifying as a statistically pertinent multivariate extension. For statisticians, distribution and quantile functions are not just probabilistic notions: above all, their empirical versions (empirical quantiles and ranks) constitute fundamental tools for inference. A multivariate extension yielding quantiles and ranks that do not match, in dimension $d \geq 2$, the properties that make traditional ranks natural and successful tools for inference in dimension one is not a statistically sound extension.

The approach we are adopting here is placing those inferential concerns at the heart of the problem.

1.1 Ranks and rank-based inference

To facilitate the exposition, let us focus on ranks and their role in testing problems. Rank-based methods naturally enter the picture in the context of semiparametric statistical models or experiments under which the distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ of some observation $\mathbf{X} = (X_1, \ldots, X_n)'$ (with real-valued X_i's), besides the finite-dimensional parameter of interest $\boldsymbol{\theta}$, also depends on the unspecified density f of some unobserved underlying residual univariate white noise, $Z_i(\boldsymbol{\theta})$, say. More precisely, assume that $\mathbf{X} \sim \mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ iff the *$\boldsymbol{\theta}$-residuals* $Z_1(\boldsymbol{\theta}), \ldots, Z_n(\boldsymbol{\theta})$ are i.i.d. with density f (although i.i.d.-*ness* can be relaxed into *exchangeability*, we will stick to i.i.d.-*ness*). In such models—call them i.i.d. noise models—testing $H^{(n)}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ (with unspecified f) reduces to the problem of testing that $Z_1(\boldsymbol{\theta}_0), \ldots, Z_n(\boldsymbol{\theta}_0)$ is i.i.d. white noise with unspecified density f. Typical examples are linear models, with $Z_i(\boldsymbol{\theta}) = X_i - \mathbf{c}'_i\boldsymbol{\theta}$ ($\mathbf{c}_i$ a q-vector of covariates, $\boldsymbol{\theta} \in \mathbb{R}^q$), or autoregressive models, with $Z_i(\theta) = X_i - \theta X_{i-1}$ (where i denotes time and $\theta \in (-1, 1)$), etc.

Invariance arguments suggest tests based on the ranks of $Z_1(\boldsymbol{\theta}_0), \ldots, Z_n(\boldsymbol{\theta}_0)$. Those tests are *distribution-free* under H_0—a finite-sample property holding in all fixed-$\boldsymbol{\theta}_0$ submodels. That distribution-freeness property is often considered as the trademark and main virtue of (univariate) ranks; it guarantees the validity of rank-based procedures (for testing $H^{(n)}_0$), irrespective of the actual density f.

Distribution-freeness (validity under unspecified f) alone is not sufficient, though, for explaining the success of ranks, and efficiency is no less important: other distribution-free methods indeed can be constructed, such as sign or runs tests, that do not perform as well (in i.i.d. noise models) as the rank-based ones. When Wilcoxon's two-sample location test (Wilcoxon 1945) was introduced on purely heuristic grounds, it was not expected to be particularly powerful. Its unexpectedly high efficiency (compared to the corresponding Student test) soon was noticed, though, and confirmed, if not explained, in Hodges and Lehmann's famous "0.864 paper" (Hodges and Lehmann 1956). Further surprising results on the power of rank-based methods came with the celebrated Chernoff and Savage (1958) result that normal-score rank-based tests, in two-sample location or regression, are uniformly more powerful (in a local asymptotic sense) than their Student competitors. Similar results have been established later on (Hallin (1994) and Hallin and Tribel (2000) in a time series context; Paindaveine (2006) and Hallin and Paindaveine (2008) in an elliptical context), where rank-based methods are shown to outperform their traditional counterparts.

A general theoretical explanation for this somewhat intriguing and unexpected efficiency of ranks was provided in Hallin and Werker (2003). In the semiparametric context of i.i.d. noise models involving some unspecified density f, indeed, the best performance one can hope for, when performing inference on the parameter of interest $\boldsymbol{\theta}$, is *semiparametric efficiency*—as developed in the classical monograph by Bickel, Klaassen, Ritov and Wellner (1993). The traditional parametric information bounds (related to the Fisher information matrices) there are replaced with semiparametric efficiency bounds which in general are strictly less favorable—the unavoidable cost of not knowing the actual f. The main result in Hallin and Werker (2003) shows that, in i.i.d. noise models, those semiparametric efficiency bounds still can be reached by means of rank-based methods. This is what we refer to as the *semiparametric efficiency preservation* property of ranks: intuitively (we refer to Hallin and Werker (2003) for a more rigorous and formal statement), this means that, in a local and asymptotic sense, all the information about the parameter of interest $\boldsymbol{\theta}$ is contained in the residual ranks, while the corresponding order statistic of residuals only contains information on the nuisance f.

Summing up, the theoretical reasons for the success of ranks for univariate statistical inference in semiparametric models are twofold:

(DF) (distribution-freeness, a validity-related exact, finite-sample property): the vector of ($\boldsymbol{\theta}$-residual) ranks is distribution-free over the (nonparametric) family $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},f} | f \in \mathcal{F}^1\}$, where $\mathcal{F}^d$ stands for the family of nonvanishing densities over $\mathbb{R}^d$ ($d \geq 1$) (see Section 2 for a more precise definition), and

(HW) (semiparametric efficiency preservation, a local and asymptotic efficiency property): the semiparametric efficiency bound (at arbitrary $(\boldsymbol{\theta}, f)$) can be reached, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$, via rank-based procedures (tests that are measurable with respect to the ranks of $\boldsymbol{\theta}$-residuals $Z_i(\boldsymbol{\theta})$).

The key property behind (HW) is the more fundamental maximal invariance property (see Section 7.1 and Chapter 6 of Lehmann and Romano (2005) for definitions and details) of ranks

(HW*) (an exact, finite-sample property) the ranks of $\boldsymbol{\theta}$-residuals are *maximal invariant* with respect to a class $\mathcal{G}^{(n)}(\boldsymbol{\theta})$ of transformations of $\mathbb{R}^n$ *generating* the fixed-$\boldsymbol{\theta}$ submodel (that is, yielding a unique orbit in the family $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},f} | f \in \mathcal{F}^1\}$ of fixed-$\boldsymbol{\theta}$ model distributions).

In Hallin and Werker (2003), the generating class $\mathcal{G}^{(n)}(\boldsymbol{\theta})$ happens to be a group—something which (see Section 7.2) will not be the case in dimension $d \geq 2$. That group structure, however, plays no role in their proofs: it is sufficient that, for any couple $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$, $\mathrm{P}^{(n)}_{\boldsymbol{\theta},h}$ of distributions in the fixed-$\boldsymbol{\theta}$ submodel, there exist a transformation in $\mathcal{G}^{(n)}(\boldsymbol{\theta})$ pushing $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ forward to $\mathrm{P}^{(n)}_{\boldsymbol{\theta},h}$.

The (unessential) restriction, in (DF), to nonvanishing densities avoids trivial problems at the boundary of bounded supports, while (HW) (unlike (HW*)) is tacitly restricted to the subset $\mathcal{F}^1_* \subset \mathcal{F}^1$ of densities f satisfying the regularity conditions (uniform local asymptotic normality, etc.) required for semiparametric efficiency to make sense. Those conditions, however, depend on the model under study; in order to avoid specifying any $\mathcal{F}^1_*$, in this paper, we focus on (HW*).

Properties (DF) and (HW) are those a statistician would like to see satisfied, with $\mathcal{F}^d$ and $\mathcal{F}^d_*$ substituted for $\mathcal{F}^1$ and $\mathcal{F}^1_*$, by the concept of ranks associated with the empirical counterpart of any sensible definition of a multivariate distribution function.

1.2 Multivariate ranks and the ordering of $\mathbb{R}^d$, $d \geq 2$

The problem of ordering $\mathbb{R}^d$ for $d \geq 2$, thus defining multivariate concepts of ranks, signs, empirical distribution functions and quantiles, is not new, and has a rather long history in statistics. Many concepts have been proposed in the literature, a complete list of which cannot be given here. Focusing again on ranks, four types of multivariate ranks, essentially, can be found:

(a) *Componentwise ranks.* The idea of componentwise ranks goes back as far as Hodges (1955), Bickel (1965) or Puri and Sen (1966, 1967, 1969). It culminates in the monograph by Puri and Sen (1971), where inference procedures based on componentwise ranks are proposed, basically, for all classical problems of multivariate analysis; more recent references are Chaudhuri and Sengupta (1993), Nordhausen, Oja, and Tyler (2006), Segers, van den Akker, and Werker (2015), ... to quote only a very few. Time-series testing methods based on the same ranks have been considered in Hallin, Ingenbleek, and Puri (1989). Componentwise ranks actually are intimately related to copula transforms, of which they constitute the empirical version: rather than solving the tricky problem of ordering $\mathbb{R}^d$, they bypass it by considering d univariate marginal rankings. As a consequence, they crucially depend on the choice of a coordinate system. Unless the underlying distribution has independent components (see Nordhausen et al. (2009), Ilmonen and Paindaveine (2011), or Hallin and Mehta (2015)), componentwise ranks in general are not even asymptotically distribution-free. Nor are they invariant under any model-generating class of transformations; a transformation-retransformation approach has been proposed by Chakraborty and Chaudhuri (1996, 1998), which ensures affine-invariance—but the group of affine transformations is not a generating group in this context. As a consequence, neither (DF), (HW*) nor (HW) are satisfied.

(b) Spatial ranks and signs. This class of multivariate ranks includes several very ingenuous, elegant and appealing concepts, proposed by several authors (Möttönen and Oja (1995); Möttönen et al. (1997); Chaudhuri (1996); Koltchinskii (1997); Oja and Randles, (2004), Oja (2010), and many others). Similar ideas have been developed by Choi and Marden (1997) and, more recently, in high dimension by Biswas, Mukhopadhyay and Ghosh (2014) and Chakraborthy and Chaudhuri (2014, 2017). We refer to Marden (1999), Oja (1999) or the monograph by Oja (2010) for a systematic exposition and exhaustive list of references. All those concepts are extending the traditional univariate ones. As a rule, however, they fail to achieve distribution-freeness (Biswas et al. (2014) is an exception, but fails on semiparametric efficiency). Their invariance properties at best extend to classes (actually, groups) of rotations, scale or affine transformations, which are not generating groups: neither (HW*) nor (HW) are satisfied.

(c) Depth-based ranks (Liu (1992), Liu and Singh (1993); He and Wang (1997); Zuo and He (2006); Zuo and Serfling (2000); ... ; see Serfling (2002, 2012) for a general introduction on statistical depth, Hallin et al. (2010) for the related concept of quantile, Zuo (2018) for a state-of-the art survey in a regression context). Depth-based ranks, in general, are distribution-free, hence satisfy (DF). At best (except for the Monge-Kantorovich depth recently proposed by Chernozhukov et al. (2017), to be considered below), they also are affine-invariant; affine transformations, however, fail to be a generating group: neither (HW*) nor (HW) hold.

(d) Mahalanobis ranks and signs/interdirections. When considered jointly with interdirections (Randles (1989)), lift interdirections (Oja and Paindaveine (2005)), Tyler angles or Mahalanobis signs (see Hallin and Paindaveine (2002a, c)), Mahalanobis ranks do satisfy both (DF) and (HW*), hence (HW), but in elliptical models only—when f is limited to the family of elliptical densities. There, they have been used, quite successfully, in a variety of multivariate models, including one-sample location (Hallin and Paindaveine 2002a), k-sample location (Um and Randles 1998), serial dependence (Hallin and Paindaveine 2002b), linear models with VARMA errors (Hallin and Paindaveine 2004a, 2005, 2006a), VAR order identification (Hallin and Paindaveine 2004b), shape (Hallin and Paindaveine 2006b; Hallin, Oja and Paindaveine 2006), homogeneity of scatter (Hallin and Paindaveine 2008), principal and common principal components (Hallin, Paindaveine and Verdebout 2010, 2013, 2014). Unfortunately, the tests developed in those references cease to be valid, and R-estimators no longer are root-n consistent, under non-elliptical densities.

None of those multivariate rank concepts, thus, is enjoying properties (DF) nor (HW)—except, but only over the class of elliptically symmetric distributions, the (pseudo)-Mahalanobis/elliptical ranks and signs. A few other concepts have been proposed as well, related to *cone orderings* (Belloni and Winkler 2011; Hamel and Kostner 2016), which require some subjective (or problem-specific) preliminary choices, and similarly fail to achieve (DF) and (HW).

The fact that, contrary to the real line $\mathbb{R}$, the real space $\mathbb{R}^d$ for $d \geq 2$ does not admit a canonical ordering places an essential difference between dimension $d = 1$ and dimensions $d \geq 2$. Whereas the same "exogenous" left-to-right ordering of $\mathbb{R}$ applies both in population and in the sample, pertinent orderings of $\mathbb{R}^d$ are bound to be "endogenous", that is, distribution-specific in populations, and data-driven (hence, random) in samples. This is the case for the concepts developed under (b)-(d) above; it also holds for the concept we are proposing in this paper. Each distribution, each sample, thus is to produce its own ordering, inducing (related forms of) quantile and distribution functions, and classes of order-preserving transformations. As a result, datasets, at best, can be expected to produce, via adequate concepts of multivariate ranks and signs, consistent empirical versions of the unavailable underlying population ordering. That consistency typically takes the form of a Glivenko-Cantelli result (GC) connecting an empirical *center-outward distribution function* to its population version. It is essential, for such a result, to hold without any moment assumptions: moment assumptions (as in Chernozhukov et al. (2017), where consistency is established under compactly supported distributions—hence under the existence of finite moments of all orders), as a rule, are inappropriate in the intrinsically ordinal context of distribution and quantile functions.

No ordering of $\mathbb{R}^d$, $d \geq 2$ moreover can be expected to be of the one-sided "left-to-right" type, since "left" and "right" do not make sense anymore. A depth-type center-outward ordering is by far more sensible. All this calls for revisiting the traditional univariate concepts from a center-outward perspective, while disentangling the population concepts from their sample counterparts.

1.3 Outline of the paper

In this paper, we show that the so-called Monge-Kantorovich ranks and signs recently proposed by Chernozhukov et al. (2017), unlike the many concepts that have been considered so far, do enjoy distribution-freeness (DF) and the maximal invariance property (HW*) which typically entails (HW). We do not go all the way, in this paper, to prove the implication from (HW*) to (HW), though: although following along the same lines, essentially, as in Hallin and Werker (2003), a formal proof indeed requires model-specific regularity assumptions, and asymptotic representation results, in the Hájek style, for the new linear rank statistics. Such results are beyond the scope and page limitations of this paper, and are the subject of ongoing work.

Using nontechnical arguments, we also show how those multivariate ranks and signs very naturally and intuitively emerge from revisiting classical univariate concepts, to which they reduce for $d = 1$. In particular, we propose a measure transportation-based concept of *center-outward distribution function*, for which we establish a Glivenko-Cantelli property in the absence of any moment assumptions. Refraining from moment assumptions calls for an approach which is entirely different from the Monge-Kantorovich optimization perspective adopted in Chernozhukov et al. (2017). The techniques considered there (and in most of the measure-transportation literature) indeed are deeply rooted in the analytical features of the Monge-Kantorovich problem, which focuses on minimizing an expected quadratic loss which, in the absence of finite second-order moments, no longer make sense. The tools we are using here are of a more fundamental geometric nature, exploiting the concept of *cyclical monotonicity* and the approach initiated by McCann (1995) (see Section 2.1 for details). This fact is emphasized by a shift in the terminology: as our approach is no longer based on Monge-Kantorovich optimization techniques, we consistently adopt the terminology *center-outward* ranks and signs for the ranks and signs associated with empirical center-outward distribution functions, despite the fact that they coincide with the Monge-Kantorovich ranks and signs introduced in Chernozhukov et al. (2017).

Section 2 provides, for those who are not familiar with measure transportation, a very succinct and elementary account of some classical facts in the area.

In Section 3, we start with revisiting the traditional concepts of univariate distribution/quantile functions and their empirical counterparts. Those traditional concepts strongly depend on the *left-to-right* nature of the canonical ordering of $\mathbb{R}$. As this *left-to-right* feature cannot be expected to extend to higher dimension, rather than the classical distribution function F, we adopt a *center-outward* form $2F - 1$, the empirical version of which naturally leads to *center-outward ranks* and *signs*. We then establish (Section 3.4), still for $d = 1$, a characterization of those center-outward distribution functions, ranks and signs in terms of measure transportation results. That characterization extends without any change to arbitrary dimension, and is exploited in Section 4 to define center-outward distribution functions, ranks and signs in $\mathbb{R}^d$.

Section 5 deals, for arbitrary d, with the Glivenko-Cantelli property of empirical center-outward distribution functions. Sections 6 and 7 study the distributional and invariance/equivariance properties of central-outward ranks and signs, establishing (DF), the independence between ranks, signs and the order statistics, and the maximal invariance property (HW*) which, as explained, leads to (HW)—hence indicating that center-outward ranks and signs fully qualify as statistically meaningful multivariate extensions of the traditional concepts, with which they coincide for $d = 1$.

1.4 Notation

Throughout, $\mathcal{F}^d$ stands for the family of nonvanishing Lebesgue densities over $\mathbb{R}^d$, $d \in \mathbb{N}$—to be precise, the family of all densities f such that, for all $D \in \mathbb{R}^+$ there exist $\Lambda_{D;f} \geq \lambda_{D;f}$ in $(0, \infty)$ such that $\lambda_{D;f} \leq f(\mathbf{x}) \leq \Lambda_{D;f}$ for $\|\mathbf{x}\| \leq D$; let $\mathcal{P}_d$ denote the corresponding family of distributions, $\mathcal{P}_d^{(n)}$ the joint distribution of i.i.d. n-tuples with common distribution in $\mathcal{P}_d$. The probability measures and distribution functions associated with densitiesf, g, ... are denoted by $\mathrm{P}_f, \mathrm{P}_g$, ..., and F, G, ..., respectively; $\mathrm{P}_f^{(n)}, \mathrm{P}_g^{(n)}, ...$ stand for the distributions of i.i.d. n-tuples with densities f, g, ... The notation $\mathbb{S}_d$, $\mathcal{S}_{d-1}$ is used for the (open) unit ball and the unit sphere in $\mathbb{R}^d$, respectively.

2 Measure transportation: Monge, Kantorovich, Brenier, McCann

Starting from a very practical problem—*How should one best move given piles of sand to fill up given holes of the same total volume?*—Gaspard Monge (1746-1818), with his 1781 *Mémoire sur la Théorie des Déblais et des Remblais*, initiated a profound mathematical theory anticipating different areas of differential geometry, linear programming, nonlinear partial differential equations, and probability.

In modern notation, the simplest and most intuitive—if not most general—formulation of Monge's problem is (in probabilistic form) as follows. Let P_1 and P_2 belong to the family $\mathcal{P}$ of probability measures over (for simplicity) $(\mathbb{R}^d, \mathcal{B}^d)$, and let $L : \mathbb{R}^{2d} \to [0, \infty]$ be a Borel-measurable loss function: $L(\mathbf{x}_1, \mathbf{x}_2)$ represents the cost of transporting $\mathbf{x}_1$ to $\mathbf{x}_2$. The objective is to find a measurable (transport) map $T_{\mathrm{P}_1;\mathrm{P}_2} : \mathbb{R}^d \to \mathbb{R}^d$ solving the minimization problem

$$\inf_T \int_{\mathbb{R}^d} L\big(\mathbf{x}, T(\mathbf{x})\big) \mathrm{dP}_1 \qquad \text{subject to} \quad T\#\mathrm{P}_1 = \mathrm{P}_2 \tag{2.1}$$

where T ranges over the set of measurable map from $\mathbb{R}^d$ to $\mathbb{R}^d$, and $T\#\mathrm{P}_1$ is the so-called *push forward of* P_1 *by* T (in statistics, a more classical but heavier notation for $T\#\mathrm{P}_1$ would be $\mathrm{P}_1^{T\mathbf{X}}$ or $\bar{T}\mathrm{P}_1$, where $\bar{T}$ is the transformation of $\mathcal{P}$ *induced by* T; see Lehmann and Romano (2005)). For simplicity, and with a slight abuse of language, we will say that T *is mapping* P_1 *to* P_2. A map $T_{\mathrm{P}_1;\mathrm{P}_2}$ achieving the infimum in (2.1) is called an *optimal transport map*, in short, an *optimal transport*, of P_1 to P_2. In the sequel, we shall restrict to the quadratic (or L^2) loss function $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$.

The problem looks simple but it is not. Monge himself (who moreover was considering the more delicate loss $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$) did not solve it, and relatively little progress was made until the 1940s, when renewed interest in the topic was triggered by the contributions of Leonid Vitalievitch Kantorovich (1912-1986; Nobel Prize in Economics in 1975) and his groundbreaking duality approach. Among the most powerful ensuing results is the *Polar Factorization Theorem* by Brenier (1987, 1991; see Chapter 3 in Villani (2003)) which implies, among other things, that for L^2 loss, if P_1 and P_2 are absolutely continuous with finite second-order moments, the solution of Monge's problem exists, is (a.e.) unique, and the gradient of a convex (potential) function—a form of multivariate monotonicity. The subject ever since has been a very active domain of mathematical analysis, with applications in various fields, from fluid mechanics to economics (see Galichon (2016)), learning, and statistics (Carlier et al. (2016); Panaretos and Zemel (2016, 2018)). It was popularized recently by the French Fields medalist Cédric Villani, with two monographs (Villani 2003, 2009), where we refer to for background reading, along with the two volumes by Rachev and Rüschendorf (1998), where the scope is somewhat closer to probabilistic and statistical concerns.

Whether described as in (2.1), or relaxed into the more general coupling form adopted by Kantorovich, the so-called Monge-Kantorovich problem remains an optimization problem, though, which only makes sense under densities for which expected costs are finite—under finite variances, thus, for quadratic loss. Such moments assumption, in a general context of distribution functions, ranks and quantiles, is not appropriate. Brenier's *Polar Factorization Theorem* relies on similar assumptions, but inspired a remarkable result by McCann (1995, page 310), hereafter the *McCann Theorem*. The nature of that theorem is geometric rather than analytical and, contrary to Monge, Kantorovitch and Brenier, does not require any moment restrictions. McCann's Theorem implies that, for any given absolutely continuous P_2, there exists, in the class of gradients of convex functions, a P_1-essentially unique element pushing P_1 forward to P_2. Under the existence of finite moments of order two, that mapping moreover coincides with the L^2-optimal (in the Monge-Kantorivich sense) transport of P_1 to P_2.

Those measure transportation results are the basis of Carlier et al. (2016)'s concept of *vector quantile regression*, and of Chernozhukov et al. (2017)'s concept of *Monge-Kantorovich depth* and related quantiles, ranks and signs; see also Ekeland et al. (2012) for precursory ideas. While Carlier et al. (2016) consider mappings to the unit cube, Chernozhukov et al. (2017) deal with mappings to general reference distributions, including the uniform over the unit ball. On the other hand, they emphasize the consistent estimation of Monge-Kantorovich depth/quantile contours, with techniques requiring compactly supported distributions (hence finite moments of all orders, which is quite regrettable when defining a quantile concept); their proofs strongly exploit Kantorovich's duality approach.

In the present paper, we privilege mappings to the uniform distribution over the unit ball, which enjoys better invariance/equivariance properties than the unit cube—the latter indeed is not unique, and possesses edges and vertices, which are "special points"—and naturally extends the elliptical case. Moreover, we are focusing on the inferential properties of quantiles, ranks and signs and, adopting McCann's geometric point of view, we manage to waive moment assumptions which, as we already stressed, are inappropriate in the context. The focus, applicability and mathematical nature of our approach, thus, is quite different from that of Chernozhukov et al. (2017).

Yet another approach is taken in a recent paper by Faugeras and Rüschendorf (2018), who propose combining a copula transform with a mapping in the Chernozhukov et al. (2017) style. This takes care of the compact support/second-order moment restriction, but results in a concept that heavily depends on the original coordinate system, which compromises the maximal invariance property (HW*) leading to (HW).

3 Distribution and quantile functions, ranks and signs in $\mathbb{R}$

The concept of empirical distribution function, hence the concepts of ranks, signs, order statistics, and quantiles, are well understood and abundantly studied in dimension one. Before introducing multivariate extensions, we therefore briefly revisit the traditional versions of those fundamental concepts and some of their main properties.

3.1 Traditional univariate concepts

Denote by $\mathbf{Z}^{(n)} := \big(Z_1^{(n)}, \ldots, Z_n^{(n)}\big)$ an n-tuple of real-valued random variables—observations or residuals associated with some parameter $\boldsymbol{\theta}$ of interest, which we emphasize, when needed, by writing $Z_i^{(n)} = Z_i^{(n)}(\boldsymbol{\theta})$. We throughout consider the case that the $Z_i^{(n)}$'s are (under parameter value $\boldsymbol{\theta}$ for the $Z_i^{(n)}(\boldsymbol{\theta})$'s) i.i.d. with density $f \in \mathcal{F}^1$, distribution P_f and distribution function F.

In dimension one, the definition of ranks is based on the canonical left-to-right ordering $\geq$ of the real line, which is not data-driven: the rank of $Z_i^{(n)}$ among $Z_1^{(n)}, \ldots, Z_n^{(n)}$ is traditionally defined as

$$R_i^{(n)} := \#\{j|\ Z_j^{(n)} \leq Z_i^{(n)}\}, \qquad i = 1, \ldots n; \tag{3.1}$$

the vector $\mathbf{R}^{(n)} := \big(R_1^{(n)}, \ldots, R_n^{(n)}\big)$ then is some random permutation of $\{1, \ldots, n\}$. Intimately related with the concept of ranks is the *dual* concept of *order statistics*, with the rth order statistic $Z_{(r)}^{(n)}$, $r = 1 \ldots, n$ implicitly defined by

$$Z^{(n)}_{(R_i^{(n)})} = Z_i^{(n)}, \qquad i = 1, \ldots, n. \tag{3.2}$$

Under the assumptions made, the vector of order statistic $\mathbf{Z}_{(\,.\,)}^{(n)} := \big(Z_{(1)}^{(n)}, \ldots, Z_{(n)}^{(n)}\big)$ is sufficient and complete, while $\mathbf{R}^{(n)}$ is uniform over the $n!$ permutations of $\{1, \ldots, n\}$, hence distribution-free. Basu's Theorem (Basu (1955); see, e.g., page 152 of Lehmann and Romano (2005)) moreover implies that $\mathbf{R}^{(n)}$ and $\mathbf{Z}_{(\,.\,)}^{(n)}$ are mutually independent.

For the empirical distribution function $F^{(n)}$, the classical definition yields

$$F^{(n)} : z \in \mathbb{R} \mapsto F^{(n)}(z) := \begin{cases} 0 & \text{if} \quad z < \min_j \big\{Z_j^{(n)}\big\} \\ \dfrac{R_i^{(n)}}{n+1} & \text{if} \quad Z_i^{(n)} = \max\big\{Z_j^{(n)}\big|\ Z_j^{(n)} \leq z\big\}; \end{cases}$$

the denominator is chosen as $(n+1)$ rather than n so that all $F^{(n)}(Z_i^{(n)})$'s take values in the open interval $(0, 1)$. The restriction $\big(F^{(n)}(Z_1^{(n)}), \ldots, F^{(n)}(Z_n^{(n)})\big)$ of $F^{(n)}$ to $\mathbf{Z}^{(n)}$ then is uniformly distributed over the $n!$ permutations of the regular grid

$$\{1/(n+1),\ 2/(n+1), \ldots, n/(n+1)\}, \tag{3.3}$$

hence distribution-free and independent of the order statistic.

The Glivenko-Cantelli Theorem tells us that

$$\sup_{z\in\mathbb{R}} \Big|F^{(n)}(z) - F(z)\Big| \longrightarrow 0 \quad \text{a.s. as } n \to \infty; \tag{3.4}$$

which, under the assumptions made (nonvanishing densities), is equivalent to the apparently weaker property (GC) that

$$\max_{1\leq i\leq n} \Big|F^{(n)}(Z_i^{(n)}) - F(Z_i^{(n)})\Big| \longrightarrow 0 \quad \text{a.s. as } n \to \infty \tag{3.5}$$

Actually, $F^{(n)}$ is entirely determined by its restriction to the observations $Z_i^{(n)}$—namely, by the n couples $(Z_i^{(n)}, F^{(n)}(Z_i^{(n)}))$, $i = 1, \ldots, n$. All other values of $F^{(n)}$ constitute an arbitrary interpolation carrying no further information—any choice of a nondecreasing interpolation would be equally legitimate and, in particular, would satisfy the same Glivenko-Cantelli property (3.4). From now on, we use the notation $F^{(n)}$ for that restriction (a data-driven mapping of the observations to the grid (3.3); any monotone nondecreasing interpolation will be denoted by $\bar{F}^{(n)}$.

3.2 Center-outward distribution and quantile function in $\mathbb{R}$

For the purpose of multidimensional generalizations, though, let us consider slightly modified concepts of distribution function, quantiles, ranks, and signs. Define the *center-outward distribution function* $\mathbf{F}_\pm$ of a distribution $\mathrm{P}_f \in \mathcal{P}_1$ as $\mathbf{F}_\pm := 2F - 1$.

Clearly, being linear transformations of each other, F and $\mathbf{F}_\pm$ carry the same information about P_f. Just as F, the center-outward distribution function $\mathbf{F}_\pm$ is a probability-integral transformation: denoting by U_1 the uniform distribution over the one-dimensional unit ball $\mathbb{S}_1 = (-1, 1)$,

$$Z \sim \mathrm{P}_f \quad \text{iff} \quad U := \mathbf{F}_\pm(Z) \sim \mathrm{U}_1. \tag{3.6}$$

Boldface is used in order to emphasize the interpretation of $\mathbf{F}_\pm$ as a vector-valued quantity: while $\|\mathbf{F}_\pm(z)\| = |2F(z) - 1|$ is the U_1-probability contents of the interval $(\pm\|\mathbf{F}_\pm(z)\|)$ (the one-dimensional ball with radius $\|\mathbf{F}_\pm(z)\|$), the unit vector $\mathbf{S}_\pm(z) := \mathbf{F}_\pm(z)/\|\mathbf{F}_\pm(z)\|$ (a point on the unit sphere $\mathcal{S}_0 = \{-1, 1\}$; $\mathbf{S}_\pm(0)$ can be defined arbitrarily) is a direction or a sign—the sign of the deviation $z - \mathrm{Med}(\mathrm{P}_f)$ of z from the median $\mathrm{Med}(\mathrm{P}_f) := F^{-1}(1/2) = \mathbf{F}_\pm^{-1}(0)$ of P_f. Those interpretations, as we shall see, will carry over to dimension $d \geq 2$.

A quantile function usually is defined as the inverse of a distribution function. Inverting $\mathbf{F}_\pm$ (which, for $\mathrm{P}_f \in \mathcal{P}_1$, is strictly increasing) yields the *center-outward quantile function*

$$\mathbf{Q}_\pm : \mathbf{u} \in \mathbb{S}_1 = (-1, 1) \mapsto \mathbf{Q}_\pm(\mathbf{u}) := \mathbf{F}_\pm^{-1}(\mathbf{u}).$$

Quantiles thus are indexed by the points $\mathbf{u}$ of the unit ball $\mathbb{S}_1 = (-1, 1)$; $\|\mathbf{u}\| \in (0, 1]$ is to be interpreted as a *quantile level*. The sets

$$\big\{\mathbf{Q}_\pm(\mathbf{u}) \bigm| \|\mathbf{u}\| = u\big\} = \big\{z_u^-, z_u^+\big\}$$

and the closed intervals

$$\big\{\mathbf{Q}_\pm(\mathbf{u}) \bigm| \|\mathbf{u}\| \leq u\big\} = \big[z_u^-, z_u^+\big]$$

where z_u^- and z_u^+ are such that $\mathrm{P}_f\big[z_u^-, \mathrm{Med}(\mathrm{P}_f)\big] = \mathrm{P}_f\big[\mathrm{Med}(\mathrm{P}_f), z_u^+\big] = u/2$ accordingly have the interpretation of *quantile contours* and *quantile regions*, at quantile level $u \in [0, 1]$.

While traditional distribution and quantile functions are associated with nested half-lines of the form $(-\infty, z_u]$ carrying probability $u \in (0, 1)$, the center-outward ones are about nested intervals $[z_u^-, z_u^+]$ (all containing $\mathrm{Med}(\mathrm{P}_f)$) with P_f-probability contents $u \in [0, 1)$, the geometry of which, unlike the traditional collection of half-lines (which is fixed), is adapted to the underlying distribution P_f. The translation of the center-outward concept in terms of the traditional one is straightforward, though, as $z_u^- = z_u$ and $z_u^+ = z_{1-u}$, where $z_\alpha := F^{-1}(\alpha)$.

3.3 Center-outward ranks and signs in $\mathbb{R}$

Turning to a sample $Z_1^{(n)}, \ldots, Z_n^{(n)}$, define the *center-outward rank* $R_{\pm;i}^{(n)}$ of $Z_i^{(n)}$ as

$$R_{\pm;i}^{(n)} := \begin{cases} \left| R_i^{(n)} - \frac{n+1}{2} \right| & \text{if } n \text{ is odd} \\ \left| R_i^{(n)} - \frac{n+1}{2} \right| + \frac{1}{2} & \text{if } n \text{ is even,} \end{cases} \tag{3.7}$$

its empirical sign as $\mathbf{S}_{\pm;i}^{(n)} := I[R_i^{(n)} > (n+1)/2] - I[R_i^{(n)} < (n+1)/2]$, and the value at $Z_i^{(n)}$ of the *empirical center-outward distribution function* as

$$\mathbf{F}_{\pm}^{(n)}(Z_i^{(n)}) := \mathbf{S}_{\pm;i}^{(n)} \frac{R_{\pm;i}^{(n)}}{\lfloor n/2 \rfloor + 1} = \begin{cases} 2F^{(n)}(Z_i^{(n)}) - 1 & \text{if } n \text{ is odd} \\ \dfrac{n+1}{n+2}\left(2F^{(n)}(Z_i^{(n)}) - 1\right) + \dfrac{1}{n+2} & \text{if } n \text{ is even,} \end{cases} \tag{3.8}$$

with values on the regular grids

$$\frac{-\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}, \ \ldots, \ \frac{-2}{\lfloor n/2 \rfloor + 1}, \frac{-1}{\lfloor n/2 \rfloor + 1}, \ 0 \ , \frac{1}{\lfloor n/2 \rfloor + 1}, \frac{2}{\lfloor n/2 \rfloor + 1}, \ \ldots, \ \frac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1} \tag{3.9}$$

(n odd), and

$$\frac{-\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1}, \ \ldots, \ \frac{-2}{\lfloor n/2 \rfloor + 1}, \frac{-1}{\lfloor n/2 \rfloor + 1}, \frac{1}{\lfloor n/2 \rfloor + 1}, \frac{2}{\lfloor n/2 \rfloor + 1}, \ \ldots, \ \frac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1} \tag{3.10}$$

(n even). Those grids are the intersection between the two unit vectors $\mathbf{u} = \pm\mathbf{1}$ and the circles with radii $1/(\lfloor n/2 \rfloor + 1)$, $2/(\lfloor n/2 \rfloor + 1)$, …, and $\lfloor n/2 \rfloor/(\lfloor n/2 \rfloor + 1)$, centered at the origin—along with the origin when n is odd.

If $Z_1^{(n)}, \ldots, Z_n^{(n)}$ are i.i.d. with some density f, the signs $\mathbf{S}_{\pm;i}^{(n)}$ are uniform over the unit sphere $\mathcal{S}_0$, and independent of the ranks $R_{\pm;i}^{(n)}$; each rank is uniformly distributed over the integers $(0, 1, 2, \ldots, \lfloor n/2 \rfloor)$ (n odd), the integers $(1, 2, \ldots, \lfloor n/2 \rfloor = n/2)$ (n even), while the n-tuple $\left(\mathbf{F}_{\pm}^{(n)}(Z_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(Z_n^{(n)})\right)$ is uniform over the $n!$ permutations of the grid (3.9) (n odd), over the $n!$ permutations of the grid (3.10) (n even).

Formula (3.8) looks complicated, but it is not: the center-outward ranks, actually, result from ordering from left to right the $\lfloor n/2 \rfloor$ observations sitting to the right of the median (with sign $\mathbf{S}_{\pm;i}^{(n)} = 1$), and ordering from right to left the $\lfloor n/2 \rfloor$ observations sitting to the left of the median (with sign $\mathbf{S}_{\pm;i}^{(n)} = -1$); the regular grids (3.9) and (3.10) on [-1,1] are replacing the traditional regular grid (3.3) of $F^{(n)}(Z_i^{(n)})$ values over [0,1].

In view of (3.8), the Glivenko-Cantelli result (3.5) for $F^{(n)}$ straightforwardly extends to $\mathbf{F}_{\pm}^{(n)}$:

$$\max_{1 \le i \le n} \left\| \mathbf{F}_{\pm}^{(n)}(Z_i^{(n)}) - \mathbf{F}_{\pm}(Z_i^{(n)}) \right\| \longrightarrow 0 \quad \text{a.s. as } n \to \infty \tag{3.11}$$

If $\mathbf{F}_{\pm}^{(n)}$ is to be defined over the whole real line, any nondecreasing interpolation $\bar{\mathbf{F}}_{\pm}^{(n)}$ of the n couples $(Z_i^{(n)}, \mathbf{F}_{\pm}^{(n)}(Z_i^{(n)}))$ provides a solution. Clearly, infinitely many choices are possible, and all of them yield a Glivenko-Cantelli statement under $\sup_{z \in \mathbb{R}}$ form (similar to (3.4)). Some are continuously differentiable, some are simply continuous (e.g., a linear interpolation), some are discontinuous, some are strictly increasing, some are step functions. Among them is the

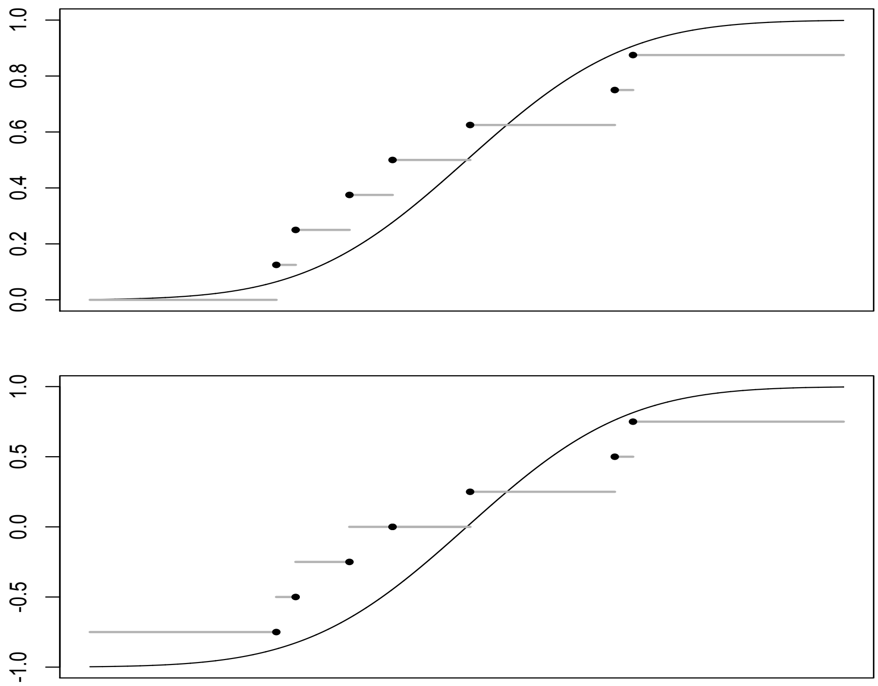

Figure 1: A classical distribution function F and its empirical counterpart $F^{(n)}$, $n = 7$ (top panel), along with (bottom panel) their center-outward versions $\mathbf{F}_{\pm}^{(n)}$ and $\mathbf{F}_{\pm}^{(n)}$, the latter with left-continuous piecewise interpolation on the left-hand side of the (empirical) median, right-continuous piecewise interpolation on the the right-hand side of the median.

continuous-from-the-left on the left-hand side of the (empirical) median, and continuous-from-the-right on the right-hand side of the median piecewise constant interpolation shown in Figure 2, which is the "most slowly center-outward increasing" one.

Clearly, the traditional ranks $R_i^{(n)}$ and the center-outward $F^{(n)}(Z_i^{(n)})$, $i = 1, \ldots, n$, generate the same σ-field: all classical rank statistics therefore can be rewritten in terms of signs and center-outward ranks. Traditional ranks, and center-outward ranks complemented with the signs, therefore, are equivalent statistics.

3.4 Relation to measure transportation

The probability-integral transformation $z \mapsto \mathbf{F}_{\pm}(z)$ from $\mathbb{R}$ to the unit ball $\mathbb{S}_1 = (-1, 1)$ is mapping the distribution $\mathrm{P} \in \mathcal{P}_1$ to the uniform distribution U_1 over $(-1, 1)$. As a monotone increasing function, it is the gradient (here, the derivative) of a convex function ψ (which is defined up to an additive constant). It follows from McCann's Theorem that it is the (essentially) unique gradient of a convex function mapping P_f to U_1. Therefore, this characterization can be adopted as the definition of $\mathbf{F}_{\pm}$. The huge advantage of this measure transportation-based definition is that it does not involve the canonical ordering of $\mathbb{R}$, and therefore readily extends to $\mathcal{P}_d$, $d \geq 2$.

4 Distribution and quantile functions, ranks and signs in $\mathbb{R}^d$

We are now ready to propose our definition of distribution and quantile functions in $\mathbb{R}^d$, along with their empirical counterparts. To start with, observe that U_1, which is the Lebesgue-uniform distribution over the unit ball $\mathbb{S}_1$, is also the product of the uniform measure over the unit sphere $\mathcal{S}_0 = \{-1, 1\}$ with a uniform measure over the unit interval of distances from the origin. We similarly define U_d as the product of the uniform measure over the unit sphere $\mathcal{S}_{d-1}$ with a uniform measure over the unit interval of distances to the origin; while we still call it *uniform over the unit ball*, U_d no longer coincides, for $d \geq 2$, with the Lebesgue-uniform measure over $\mathbb{S}_d$.

4.1 Center-outward distribution and quantile functions in $\mathbb{R}^d$

Before turning to the definition of center-outward distribution and quantile functions in $\mathbb{R}^d$, we need the following property, which guarantees the existence, uniqueness and continuity of the concepts, and is borrowed, with some minor modifications, from Theorem 1.1 in Figalli (2018).

Proposition 4.1 *Let* $\mathrm{P} \in \mathcal{P}_d$. *Then,*

(i) the gradient of convex function $\nabla\Psi$ *pushing* P *forward to the uniform* U_d *over the unit ball* $\mathbb{S}_d$ *is unique; the set* $K := \{\mathbf{x} | \nabla\Psi(\mathbf{x}) = \mathbf{0}\}$ *is compact and has Lebesgue measure zero;*

(ii) the restriction of $\nabla\Psi$ *to* $\mathbb{R}^d \setminus K$ *is a homeomorphism from* $\mathbb{R}^d \setminus K$ *to* $\mathbb{S}_d \setminus \{\mathbf{0}\}$*, with inverse (defined on* $\mathbb{S}_d \setminus \{\mathbf{0}\}$*)* $\nabla\Psi^*$*, where* Ψ^* *is the Legendre transform of* Ψ*; for* $d = 1, 2$*, however,* K *consists of a single point, and* $\nabla\Psi$ *is a homeomorphism from* $\mathbb{R}^d$ *to* $\mathbb{S}_d$*;*

(iii) if P *has Lebesgue density* f*,* $f(\mathbf{x}) = c_d^{-1} H_{\Psi^*}(\mathbf{x}) \|\nabla\Psi(\mathbf{x})\|^{1-d}$*,* $\mathbf{x} \in \mathbb{R}^d \setminus K$*, where the norming constant* $c_d := 2\pi^{d/2}/\Gamma(d/2)$ *is the area of the unit sphere* $\mathcal{S}_{d-1}$ *and* H_{Ψ^*} *the Hessian of* Ψ^**.*

The following definitions then coincide, for $d = 1$, with the univariate ones given in Section 3.2.

Definition 4.1 *Let* $\mathrm{P} \in \mathcal{P}_d$. *The (center-outward)* distribution function $\mathbf{F}_\pm$ *of* P *is the unique gradient of a convex function mapping* $\mathbb{R}^d$ *to the open unit ball* $\mathbb{S}_d$ *and pushing* P *forward to the uniform* U_d *over* $\mathbb{S}_d$. *The corresponding (center-outward)* quantile function *is* $\mathbf{Q}_\pm := \mathbf{F}_\pm^{-1}$. *Denoting by* $q\bar{\mathbb{S}}_d$ *and* $q\mathcal{S}_{d-1}$ *the closed ball and the hypersphere with radius* $q \in (0, 1)$ *centered at the origin, the quantile function* $\mathbf{Q}_\pm$ *characterizes* quantile regions $\mathbb{C}(q) := \mathbf{Q}_\pm(q\bar{\mathbb{S}}_d)$ *and* quantile contours $\mathcal{C}(q) := \mathbf{Q}_\pm(q\mathcal{S}_{d-1})$*, respectively, of order* q*; the elements of* $\mathbb{C}(0) = \mathcal{C}(0) = \mathbf{Q}_\pm(\mathbf{0})$ *(a compact set with Lebesgue measure zero) are called* center-outward medians*.*

The following elementary properties of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ readily follow from the definition, or are immediate consequences of Proposition 4.1; details are left to the reader.

Proposition 4.2 *Let* P *have a density* $f \in \mathcal{F}^d$. *Then,*

(i) $\mathbf{F}_\pm$ *is a probability integral transformation of* $\mathbb{R}^d$*: namely,* $\mathbf{Z} \sim \mathrm{P}$ *iff* $\mathbf{F}_\pm(\mathbf{Z}) \sim \mathrm{U}_d$*;*

(ii) for $d = 1, 2$*,* $\mathbf{F}_\pm$ *and* $\mathbf{Q}_\pm$ *are homeomorphisms between* $\mathbb{R}^d$ *and* $\mathbb{S}_d$*, respectively, and the center-outward median* $\mathbf{Q}_\pm(\mathbf{0})$ *is uniquely defined; for* $d \geq 3$*, the restrictions of* $\mathbf{F}_\pm$ *and* $\mathbf{Q}_\pm$ *to* $\mathbb{R}^d \setminus \mathbf{Q}_\pm(\mathbf{0})$ *and* $\mathbb{S}_d \setminus \{\mathbf{0}\}$ *are homeomorphisms between* $\mathbb{R}^d \setminus \mathbf{Q}_\pm(\mathbf{0})$ *and* $\mathbb{S}_d \setminus \{\mathbf{0}\}$*, and the center-outward medians form a compact set of measure zero;*

(iii) the quantile regions $\mathbb{C}(q)$, with boundaries $\mathcal{C}(q)$, are connected, compact, and nested as q increases from 0 to 1; their probability contents is q, $q \in [0, 1)$.

The center-outward distribution and quantile functions $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ thus preserve the probability integral transformation nature of univariate distribution functions, and the interpretation of univariate quantile contours as the boundaries . The terminology *quantile region* and *quantile contour* of order q is justified (for $\mathrm{P} \in \mathcal{P}_d$) by (iii).

For any given distribution $\mathrm{P} \in \mathcal{P}_d$, $\mathbf{F}_\pm$ induces a (partial) ordering of $\mathbb{R}^d$ similar to the ordering induced on the unit ball by the system of polar coordinates, and actually coincides with the "vector rank transformation" considered in Chernozhukov et al. (2017); the compact support and Cafarelli assumptions made there are not needed here, though. The quantile contours $\mathcal{C}(q)$ also have the interpretation of depth contours associated with the Monge-Kantorovich depth concept considered in the same reference.

4.2 Center-outward ranks and signs in $\mathbb{R}^d$

Turning to the sample situation, let $\mathbf{Z}^{(n)} := \left(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}\right)$ denote an n-tuple of random vectors— observations or residuals associated with some parameter $\boldsymbol{\theta}$ of interest. We throughout consider the case that the $\mathbf{Z}_i^{(n)}$'s are (possibly, under parameter value $\boldsymbol{\theta}$) i.i.d. with density $f \in \mathcal{F}^d$, distribution P and center-outward distribution function $\mathbf{F}_\pm$.

For the empirical counterpart $\mathbf{F}_\pm^{(n)}$ of $\mathbf{F}_\pm$, we propose the following extension of the univariate concept described in Section 3.3. Assuming $d \geq 2$, let n factorize into

$$n = n_R n_S + n_0, \qquad n_R,\ n_S,\ n_0 \in \mathbb{N}, \quad 0 \leq n_0 < \min(n_R, n_S) \tag{4.1}$$

where $n_R \to \infty$ and $n_S \to \infty$ as $n \to \infty$; (4.1) is extending to $d \geq 2$ the factorization of n into $n = \lfloor \frac{n}{2} \rfloor 2 + n_0$ with $n_0 = 0$ (n even) or $n_0 = 1$ (n odd) that leads, for $d = 1$, to the grids (3.9) and (3.10).

Next, consider a sequence of "regular grids" of $n_R n_S$ points in the unit ball $\mathbb{S}_d$ obtained as the intersection between

- a "regular" n_S-tuple $(\mathbf{u}_1, \ldots \mathbf{u}_{n_S})$ of unit vectors, and
- the n_R hyperspheres centered at the origin, with radii $\dfrac{1}{n_R + 1}, \dfrac{2}{n_R + 1}, \ \ldots, \ \dfrac{n_R}{n_R + 1}$,

along with n_0 copies of the origin whenever $n_0 > 0$. In theory, by a "regular" n_S-tuple $(\mathbf{u}_1, \ldots \mathbf{u}_{n_S})$, we only mean that the sequence of uniform discrete distributions over $\{\mathbf{u}_1, \ldots \mathbf{u}_{n_S}\}$ converges weakly, as $n_S \to \infty$, to the uniform distribution over $\mathcal{S}_{d-1}$. In practice, each n_S-tuple should be "as uniform as possible". For $d = 2$, perfect regularity can be achieved by dividing the unit circle into n_S arcs of equal length $2\pi/n_S$. Starting with $d = 3$, however, this typically is no longer possible. A random array of n_S independent and uniformly distributed unit vectors does satisfy (almost surely) the weak convergence requirement. More regular deterministic arrays (with faster convergence) can be considered, though, such as the *low-discrepancy sequences* of the type considered in numerical integration and Monte-Carlo methods (see, e.g., Niederreiter (1992), Judd (1998), Dick and Pillichshammer (2014), or Santner et al. (2003)), which are current practice in numerical integration and the design of computer experiments.

The resulting grid of $n_R n_S$ points then is such that the discrete distribution with probability masses $1/n$ at each gridpoint and probability mass n_0/n at the origin—call it *uniform over the* *augmented grid*—converges weakly to the uniform U_d over the ball $\mathbb{S}_d$—recall that, by uniform, we mean the product of a uniform over $\mathcal{S}_{d-1}$ (the distribution of a multivariate sign) and a uniform over the unit radius (the distribution of a distance to the origin). That grid, along with the n_0 copies of the origin, is called the *augmented grid* (n points).

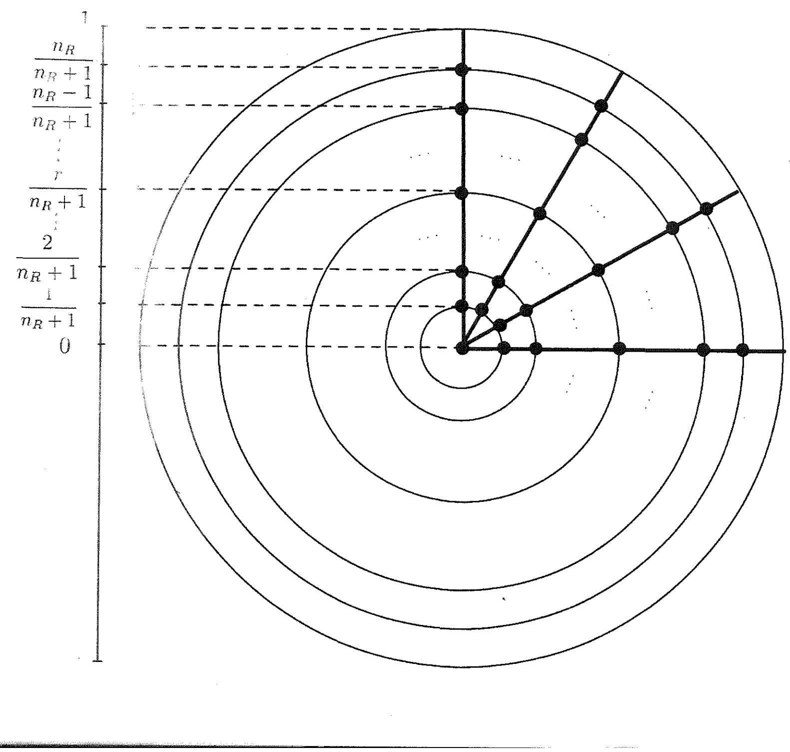

Figure 2: A regular grid of $n = n_R n_s$ points over $\mathbb{S}_2$.

We then define $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})$, $i = 1, \ldots, n$ as the solution of an optimal coupling problem between the observations and the augmented grid. Let $\mathcal{T}$ denote the set of all possible bijective mappings between $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ and the n points of the augmented grid just described. Under the assumption made, the $\mathbf{Z}_i^{(n)}$'s are all distinct with probability one, so that $\mathcal{T}$ contains $n!/n_0!$ classes of $n_0!$ indistinguishable couplings each (two couplings T_1 and T_2 are indistinguishable if $T_1(\mathbf{Z}_i^{(n)}) = T_2(\mathbf{Z}_i^{(n)})$ for all i).

Definition 4.2 *The empirical center-outward distribution function is the (random) mapping*

$$\mathbf{F}_{\pm}^{(n)} : \mathbf{Z}^{(n)} := \left(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}\right) \mapsto \left(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})\right)$$

satisfying

$$\sum_{i=1}^{n} \left\|\mathbf{Z}_i^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\right\|^2 = \min_{T \in \mathcal{T}} \sum_{i=1}^{n} \left\|\mathbf{Z}_i^{(n)} - T(\mathbf{Z}_i^{(n)})\right\|^2 \tag{4.2}$$

or, equivalently,

$$\sum_{i=1}^{n} \left\|\mathbf{Z}_i^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\right\|^2 = \min_{\pi} \sum_{i=1}^{n} \left\|\mathbf{Z}_{\pi(i)}^{(n)} - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\right\|^2 \tag{4.3}$$

where the set $\{\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})|\ i = 1, \dots, n\}$ *coincides with the* n *points of the augmented grid and* π *ranges over the* $n!$ *possible permutations of* $\{1, 2, \dots, n\}$.

The n-tuple

$$\big\{\big(\mathbf{Z}_1^{(n)}, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_1^{(n)})\big), \dots, \big(\mathbf{Z}_n^{(n)}, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_n^{(n)})\big)\big\} \tag{4.4}$$

is thus any of the ($n_0!$ with probability one) indistinguishable couplings between the n observations and the n points of the augmented grid that minimize, over the $n!$ possible couplings, the sum (the mean) of within-pairs squared distances—a trivial and purely formal multiplicity that does not occur for $n_0 = 0$ or 1. Determining such a coupling is a standard optimal assignment problem, which clearly takes the form of a linear program for which efficient operations research algorithms are available.

Reinterpreting (4.2)-(4.3) as a (conditional on the sample) expected transportation cost, the same optimal coupling(s) also constitute(s) the optimal L^2 transport mapping the sample empirical distribution to the uniform discrete distribution over the augmented grid (and, conversely, the two problems being entirely symmetric, the optimal L^2 transport mapping the uniform discrete distribution over the augmented grid to the sample empirical distribution). Classical results (see, again, McCann (1995)) then show that optimality is achieved (that is, (4.2)-(4.3) is satisfied) iff the so-called *cyclical monotonicity* property holds for the n-tuple (4.4). Except for a set N_0^{nd} with Lebesgue measure zero in $\mathbb{R}^{nd}$ (those points for which the minimal distance, in (4.2)-(4.3), is the same for at least two permutations of the grid—a finite collection of linear subspaces with dimension less than nd), and apart from the trivial multiplicity just mentioned, the solution is unique.

Definition 4.3 *A subset* S *of* $\mathbb{R}^d \times \mathbb{R}^d$ *is said to be* cyclically monotone *if, for any finite collection of points* $\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_k, \mathbf{y}_k)\} \subseteq S$,

$$\langle \mathbf{y}_1,\ \mathbf{x}_2 - \mathbf{x}_1\rangle + \langle \mathbf{y}_2,\ \mathbf{x}_3 - \mathbf{x}_2\rangle + \ldots + \langle \mathbf{y}_k,\ \mathbf{x}_1 - \mathbf{x}_k\rangle \leq 0. \tag{4.5}$$

The subdifferential of of a convex function does enjoy cyclical monotonicity, which heuristically can be interpreted as a discrete version of the fact that a smooth convex function has a positive semi-definite second-order differential.

Note that a finite subset $S = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone iff (4.5) holds for $k = n$—equivalently, iff, among all pairings of $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $(\mathbf{y}_1, \dots, \mathbf{y}_n)$, S maximizes $\sum_{i=1}^n \langle \mathbf{x}_i, \mathbf{y}_i\rangle$ (that is, maximizes an empirical correlation), or minimizes $\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{x}_i\|^2$ (an empirical distance). In other words, a finite subset S is cyclically monotone iff the couples $(\mathbf{x}_i, \mathbf{y}_i)$ are a solution of the optimal assignment problem with assignment cost $\|\mathbf{y}_i - \mathbf{x}_i\|^2$. The L^2 transportation cost considered here is thus closely related to the concept of convexity and the geometric property of cyclical monotonicity; it does not play the statistical role of an estimation loss function—the L^2 distance between the empirical transport and its population counterpart is never considered—its expectation anyway could be infinite.

Associated with our definition of an empirical center-outward distribution function $\mathbf{F}_\pm^{(n)}$ are the following concepts of

- *center-outward ranks* $R_{\pm,i}^{(n)} := (n_R + 1)\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|$,
- *center-outward signs* $\mathbf{S}_{\pm,i}^{(n)} := \mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})/\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|$ (if $\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)}) = \mathbf{0}$, put $\mathbf{S}_{\pm,i}^{(n)} := \mathbf{0}$),
- *center-outward quantile contours* $\mathcal{C}_{\pm;\mathbf{Z}^{(n)}}^{(n)}(j/n_R) := \{\mathbf{Z}_i^{(n)}|\ R_{\pm,i}^{(n)} = j/(n_R + 1)\}$, and

– *center-outward quantile regions* $\mathbb{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}(j/n_R) := \{\mathbf{Z}^{(n)}_i |\, R^{(n)}_{\pm,i} \leq j/(n_R+1)\}$,

$i = 0, 1, \dots, n$, $j = 0, 1, \dots, n_R$; j/n_R is an empirical probability contents, to be interpreted as a quantile order. The center-outward quantile contours and regions defined here are finite collections of observed points; the problem of turning them into continuous contours enclosing compact regions is treated in del Barrio et al. (2018).

Up to this point, we have defined multivariate generalizations of the univariate concepts of center-outward distribution and quantile functions, center-outward ranks and signs, all reducing to their univariate analogues in case $d = 1$. However, it remains to show that those extensions are adequate in the sense that they enjoy in $\mathbb{R}^d$ the strong properties that make the success of their univariate counterparts—namely,

(GC) a Glivenko-Cantelli-type asymptotic relation between $\mathbf{F}^{(n)}_{\pm}$ and $\mathbf{F}_{\pm}$,

(DF) finite-n distribution-freeness (with respect to $f \in \mathcal{F}^d$), and

(HW*) the maximal invariance property leading to semiparametric preservation.

Establishing those three properties is the objective of Sections 5, 6, and 7, respectively.

5 Glivenko-Cantelli

With the definitions adopted in Section 4, the traditional Glivenko-Cantelli theorem, under its center-outward form (3.11), holds, essentially *ne varietur*, in $\mathbb{R}^d$.

Proposition 5.1 *Let* $\mathbf{Z}^{(n)}_i, \dots, \mathbf{Z}^{(n)}_i$ *be i.i.d. with distribution* $\mathrm{P} \in \mathcal{P}_d$. *Then,*

$$\max_{1 \leq i \leq n} \left\| \mathbf{F}^{(n)}_{\pm}(\mathbf{Z}^{(n)}_i) - \mathbf{F}_{\pm}(\mathbf{Z}^{(n)}_i) \right\| \longrightarrow 0 \quad \text{a.s. as } n \to \infty. \tag{5.1}$$

This proposition considerably reinforces, under more general assumptions (no second-order moments), an early strong consistency result by Cuesta-Albertos et al. (1997). The proof of (5.1) is postponed to Section 8.2.

Section 4 so far only provides a definition of $\mathbf{F}^{(n)}_{\pm}$ computed at the sample values $\mathbf{Z}^{(n)}_i$. If $\mathbf{F}^{(n)}_{\pm}$ is to be defined at all $\mathbf{z} \in \mathbb{R}^d$, an interpolation $\bar{\mathbf{F}}^{(n)}_{\pm}$, similar for instance to the one shown, for $d = 1$, in Figure 2, has to be constructed. Such interpolation should belong to the class of gradients of convex functions from $\mathbb{R}^d$ to $\mathbb{S}_d$, so that the resulting contours (the curves or hypersurfaces with equation $\|\bar{\mathbf{F}}^{(n)}_{\pm}(\mathbf{z})\| = \|\mathbf{F}^{(n)}_{\pm}(\mathbf{Z}^{(n)}_i)\|$ for some i) have the nature of (continuous) quantile contours. Moreover, they still should enjoy (now under a $\sup_{\mathbf{z} \in \mathbb{R}^d}$ form similar to (3.4)) the Glivenko-Cantelli strong consistency property. Constructing such interpolations is considerably more delicate for $d \geq 2$ than in the univariate case, and is the subject of the companion paper by del Barrio et al. (2018), where we also refer to for numerical implementation and pictures. Here, we restrict to the $\max_{1 \leq i \leq n}$ form (5.1) of Glivenko-Cantelli. It should be insisted, though, that this limitation is not really restrictive, as such interpolations do not bring any additional information on the population distribution, and are mainly intended for a graphical depiction of contours (in dimension $d \leq 3$, thus).

Proposition 5.1 has an important corollary in the case of elliptical densities. Recall that a d-dimensional random vector $\mathbf{X}$ has elliptical distribution $\mathrm{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma},f}$ with location $\boldsymbol{\mu} \in \mathbb{R}^d$, positive definite symmetric $d \times d$ scatter matrix $\mathbf{\Sigma}$ and radial density f iff $\mathbf{Z} := \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has spherical distribution $\mathrm{P}_{\mathbf{0},\boldsymbol{I},f}$, which holds iff

$$\mathbf{F}_{\text{ell}}(\mathbf{Z}) := \frac{\mathbf{Z}}{\|\mathbf{Z}\|} F\big(\|\mathbf{Z}\|\big) \sim \mathrm{U}_d, \tag{5.2}$$

where F, with density f, is the distribution function of $\|\mathbf{Z}\|$.

The mapping $\mathbf{Z} \mapsto \mathbf{F}_{\text{ell}}(\mathbf{Z})$ is thus a probability-integral transformation; it is shown in Chernozhukov et al. (2017) (Section 2.4) that it actually coincides with $\mathbf{Z}$'s center-outward distribution function $\mathbf{F}_{\pm}$. Let $\mathbf{X}_i^{(n)}$, $i = 1, \ldots, n$ be an i.i.d. n-tuple: the empirical version of $\mathbf{F}_{\text{ell}}$, based on Mahalanobis ranks (the ranks $R_i^{(n)}$ of the estimated residuals $\mathbf{Z}_i^{(n)} := \hat{\mathbf{\Sigma}}^{(n)-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}^{(n)})$) and Mahalanobis signs (the unit vectors $\mathbf{U}_i^{(n)} := \hat{\mathbf{\Sigma}}^{(n)-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}^{(n)})/\|\hat{\mathbf{\Sigma}}^{(n)-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}^{(n)})\|$) is, for the ith observation,

$$\mathbf{F}_{\text{ell}}^{(n)}(\mathbf{Z}_i^{(n)}) := \frac{R_i^{(n)}}{n+1} \mathbf{U}_i^{(n)},$$

where $\hat{\boldsymbol{\mu}}^{(n)}$ and $\hat{\mathbf{\Sigma}}^{(n)}$ are, when $\mathbf{X}_i^{(n)}$, $i = 1, \ldots, n$ are i.i.d. with elliptical distribution $\mathrm{P}_{\boldsymbol{\mu},\mathbf{\Sigma},f}$, symmetric and weakly consistent estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, respectively.

Proposition 5.2 *Let Let $\mathbf{X}_i^{(n)}$, $i = 1, \ldots, n$ be an i.i.d. with elliptical distribution $\mathrm{P}_{\boldsymbol{\mu},\mathbf{\Sigma},f}$, and assume that $\hat{\boldsymbol{\mu}}^{(n)}$ and $\hat{\mathbf{\Sigma}}^{(n)}$ are strongly consistent estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, respectively. Then, $\mathbf{F}_{ell}$ and $\mathbf{F}_{\pm}$ coincide, and*

$$\max_{1 \le i \le n} \|\mathbf{F}_{ell}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\|, \quad \textit{hence also} \quad \max_{1 \le i \le n} \|\mathbf{F}_{ell}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|$$

tend to zero a.s., as $n \to \infty$, where $\mathbf{F}_{ell}$ is given in (5.2) and $\mathbf{F}_{\pm}$ denotes the center-outward distribution function of $\mathrm{P}_{\mathbf{0},\boldsymbol{I},f}$.

This result connects the center-outward ranks and signs with the well-studied elliptical ranks and signs. The consistency of $\mathbf{F}_{\text{ell}}^{(n)}$, however, only holds under the assumption of ellipticity, whereas $\mathbf{F}_{\pm}^{(n)}$ remains consistent under any density $f \in \mathcal{F}^d$. Note also that $\mathbf{F}_{\text{ell}}^{(n)}$ determines n ellipsoidal contours (n distinct values for the $\mathbf{F}_{\text{ell}}^{(n)}(\mathbf{Z}_i^{(n)})$'s), while $\mathbf{F}_{\pm}^{(n)}$ only determines n_R of them (which for finite n do not define an ellipsoid). See Section 8.2 for a proof.

6 Distribution-freeness

Call *order statistic* the un-ordered n-tuple $\mathbf{Z}^{(n)}$—equivalently, an arbitrarily ordered version of the same, such as $\mathbf{Z}_{(\cdot)}^{(n)} := \big(\mathbf{Z}_{(1)}^{(n)}, \ldots, \mathbf{Z}_{(n)}^{(n)}\big)$, where $\mathbf{Z}_{(i)}^{(n)}$ is such that its first component is the ith order statistic of the n-tuple of first components. The following result extends to the center-outward case the usual finite-sample distributional properties of the order statistic and the vector of ranks; see Section 8.3 for a proof.

Proposition 6.1 *Let $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d$, center-outward distribution function $\mathbf{F}_{\pm}$, and empirical center-outward distribution function $\mathbf{F}_{\pm}^{(n)}$. Then,*

(i) the order statistic $\mathbf{Z}_{(\cdot)}^{(n)}$ is sufficient and complete,

(ii) (DF) $\big(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})\big)$ *is uniformly distributed over the* $n!/n_0!$ *permutations with repetitions of the augmented grid described in Section 4.2 with the origin counted as* n_0 *indistinguishable points,*

(iii) for $n_0 = 0$ *or 1, the vector of center-outward ranks* $\big(R_{\pm,1}^{(n)}, \ldots, R_{\pm,n}^{(n)}\big)$ *and the vector of center-outward signs* $\big(\mathbf{S}_{\pm,1}^{(n)}, \ldots, \mathbf{S}_{\pm,n}^{(n)}\big)$ *are mutually independent, and*

(iv) $\mathbf{Z}_{(\cdot)}^{(n)}$ *and* $\big(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \ldots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)})\big)$ *are mutually* P*-independent.*

In (iii), we assume $n_0 = 0$ or 1; for other values of n_0, a mild dependence is induced by the fact that the origin plays an n_0-tuple role in the matching between the observations and the grid; that dependence rapidly fades away as n increases. Also,note that, still for $n_0 = 0$ or 1 (hence, $n_0! = 1$), the the $n!/n_0!$ permutations with repetitions in (ii) reduce to the $n!$ "ordinary" permutations of the gridpoint.

7 Invariance, equivariance, and semiparametric efficiency

7.1 Maximal invariance and semiparametric efficiency

Hallin and Werker (2003) have established for $d = 1$ the close connection between semiparametric efficiency and maximal invariance—a connection that explains the good performances of rank-based inference in semiparametric models where the infinite-dimensional nuisance is the unspecified density of some unobserved underlying white noise.

More precisely, denote by $\mathbf{X}^{(n)} = (\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)})$ the observation described by the semiparametric model $\mathcal{P}^{(n)} = \{\mathrm{P}_{\boldsymbol{\theta};f}^{(n)} |\ \boldsymbol{\theta} \in \mathbb{R}^k,\ f \in \mathcal{F}^d\}$ where $\boldsymbol{\theta} \in \mathbb{R}^k$ is some Euclidean parameter of interest and $f \in \mathcal{F}^d$ some unspecified white noise density. Assume, for each $\boldsymbol{\theta}$, the existence of an invertible *residual function* $\mathbf{Z}^{(n)}(\boldsymbol{\theta}) : \mathbf{x}^{(n)} \in \mathbb{R}^{nd} \mapsto \mathbf{Z}^{(n)}(\mathbf{x}^{(n)}; \boldsymbol{\theta}) = \big(\mathbf{Z}_1(\boldsymbol{\theta}), \ldots, \mathbf{Z}_n(\boldsymbol{\theta})\big)$, with inverse $\mathbf{Z}_{\leftarrow}^{(n)}(\boldsymbol{\theta})$, such that, letting $\mathbf{Z}_i^{(n)}(\boldsymbol{\theta}) := \mathbf{Z}_i(\mathbf{X}^{(n)}; \boldsymbol{\theta})$,

$$\mathbf{X}^{(n)} \sim \mathrm{P}_{\boldsymbol{\theta};f}^{(n)} \quad \text{iff} \quad \mathbf{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \mathbf{Z}_n^{(n)}(\boldsymbol{\theta}) \ \text{ i.i.d. with density } f.$$

Assume that the (parametric) fixed-f submodels $\mathcal{P}_f^{(n)} := \{\mathrm{P}_{\boldsymbol{\theta};f}^{(n)} |\ \boldsymbol{\theta} \in \mathbb{R}^k\},\ \ f \in \mathcal{F}^d$ are locally asymptotically normal with central sequences $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$, and regular enough for semiparametric efficiency in the manner of Bickel et al. (1993) to make sense (which, in general, requires restricting $\mathcal{F}^d$ to some subfamily $\mathcal{F}_*^d$ in a way that depends on the model under study). Denote by $\boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\theta})$ the corresponding semiparametrically efficient central sequences (the projection of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ along the so-called *tangent spaces*). Assume moreover that, for any $\boldsymbol{\theta}$, the (nonparametric) fixed-$\boldsymbol{\theta}$ submodel $\mathcal{P}_{\boldsymbol{\theta}}^{(n)} := \{\mathrm{P}_{\boldsymbol{\theta};f}^{(n)} |\ f \in \mathcal{F}^d\}$, $\boldsymbol{\theta} \in \mathbb{R}^k$ is generated by a class of transformations $\mathcal{G}_{\boldsymbol{\theta}}^{(n)} = \{\mathfrak{g}_{\boldsymbol{\theta}}^{(n)}\}$ acting on $\mathbb{R}^{nd}$, admitting $\mathbf{T}_{\boldsymbol{\theta}}^{(n)}(\mathbf{X}^{(n)})$ as maximal invariant—that is, assume that, for any f, $g \in \mathcal{F}^d$, there exists $\mathfrak{g}_{\boldsymbol{\theta};f,g}^{(n)} \in \mathcal{G}_{\boldsymbol{\theta}}^{(n)}$ and $\mathbf{T}_{\boldsymbol{\theta}}^{(n)}$ such that

$$\mathrm{P}_{\boldsymbol{\theta};g}^{(n)} = \mathfrak{g}_{\boldsymbol{\theta};f,g}^{(n)} \# \mathrm{P}_{\boldsymbol{\theta};f}^{(n)}, \tag{7.1}$$

and, for any $\mathbf{x}^{(n)}$, $\mathbf{y}^{(n)} \in \mathbb{R}^{nd}$, there exists $\mathfrak{g}_{\boldsymbol{\theta}}^{(n)} \in \mathcal{G}_{\boldsymbol{\theta}}^{(n)}$ such that

$$\mathbf{y}^{(n)} = \mathfrak{g}_{\boldsymbol{\theta}}^{(n)} \mathbf{x}^{(n)} \quad \text{iff} \quad \mathbf{T}_{\boldsymbol{\theta}}^{(n)}(\mathbf{y}^{(n)}) = \mathbf{T}_{\boldsymbol{\theta}}^{(n)}(\mathbf{x}^{(n)}). \tag{7.2}$$

Then, typically (Hallin and Werker 2003), under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, as $n \to \infty$,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}) := \mathrm{E}\Big[\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) \big|\ \mathbf{T}^{(n)}_{\boldsymbol{\theta}}(\mathbf{X}^{(n)})\Big] = \boldsymbol{\Delta}^{(n)*}_f(\boldsymbol{\theta}) + o_{\mathrm{P}}(1). \tag{7.3}$$

.

The huge advantage of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta})$ over $\boldsymbol{\Delta}^{(n)*}_f(\boldsymbol{\theta})$ is that, being measurable with respect to the maximal invariant $\mathbf{T}^{(n)}_{\boldsymbol{\theta}}$ of a generating family of transformations, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta})$, contrary to $\boldsymbol{\Delta}^{(n)*}_f(\boldsymbol{\theta})$, is distribution-free under $\mathcal{P}^{(n)}_{\boldsymbol{\theta}}$: its distribution under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};g}$ does not depend on $g \in \mathcal{F}^d$. As a consequence, the semiparametric efficiency bounds associated with some reference density f can be attained by distribution-free $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta})$-based, hence $\mathbf{T}^{(n)}_{\boldsymbol{\theta}}$-measurable, inference—the validity of which holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};g}$ for any $g \in \mathcal{F}^d$. Call this the *semiparametric efficiency preservation* property of $\mathbf{T}^{(n)}_{\boldsymbol{\theta}}$.

Typically, the family $\mathcal{G}^{(n)}_{\boldsymbol{\theta}}$ is the residual-transformation-retransformation version

$$\mathbf{Z}^{(n)}_{\leftarrow}(\boldsymbol{\theta}) \circ \mathcal{G}^{(n)} \circ \mathbf{Z}^{(n)}(\boldsymbol{\theta}) = \Big\{\mathfrak{g}^{(n)}_{\boldsymbol{\theta}} = \mathbf{Z}^{(n)}_{\leftarrow}(\boldsymbol{\theta}) \circ \mathfrak{g}^{(n)} \circ \mathbf{Z}^{(n)}(\boldsymbol{\theta}) \big|\ \mathfrak{g}^{(n)} \in \mathcal{G}^{(n)}\Big\}$$

of some family $\mathcal{G}^{(n)} := \{\mathfrak{g}^{(n)}\}$ of transformations acting on the residuals $\mathbf{Z}^{(n)}(\boldsymbol{\theta})$, with maximal invariant $\mathbf{T}^{(n)}\big(\mathbf{Z}^{(n)}(\boldsymbol{\theta})\big)$. Equation (7.1) then takes the form (for all f, $g \in \mathcal{F}^d$)

$$\mathrm{P}^{(n)}_g = \mathfrak{g}^{(n)}_{f,g} \# \mathrm{P}^{(n)}_f \quad \text{for some } \mathfrak{g}^{(n)}_{f,g} \in \mathcal{G}^{(n)}. \tag{7.4}$$

Denoting by $\mathfrak{g}^{\otimes n}$ a transformation of $\mathbb{R}^{nd}$ factorizing as $\mathfrak{g}^{\otimes n} : (\mathbf{z}_1, \ldots, \mathbf{z}_n) \in \mathbb{R}^{nd} \mapsto (\mathfrak{g}\mathbf{z}_1, \ldots, \mathfrak{g}\mathbf{z}_n)$ where $\mathfrak{g}$ is acting on $\mathbb{R}^d$, this type of result holds, when $d = 1$, with

$$\mathcal{G}^{(n)} = \big\{\mathfrak{g}^{(n)} = \mathfrak{g}^{\otimes n} \big|\ \mathfrak{g} : \mathbb{R} \to \mathbb{R} \text{ monotone increasing continuous, } \lim_{z \to \pm\infty} \mathfrak{g}(z) = \pm\infty\big\} \tag{7.5}$$

and $\mathfrak{g}^{\otimes n}_{f,g} = \big(G^{-1} \circ F\big)^{\otimes n}$. A maximal invariant is the vector $\mathbf{T}^{(n)} = \big(R^{(n)}_1, \ldots, R^{(n)}_n\big)$ of (residual) ranks, where $R^{(n)}_i$ is the rank of $Z^{(n)}_i$ among $Z^{(n)}_1, \ldots, Z^{(n)}_n$—equivalently, the center-outward n-tuple $\big(\mathbf{F}^{(n)}_{\pm}(Z^{(n)}_1), \ldots, \mathbf{F}^{(n)}_{\pm}(Z^{(n)}_n)$; see Hallin and Werker (2003) or Hallin and La Vecchia (2018) for examples.

In the sequel, for the sake of simplicity, we are dropping the nature of $\mathbf{Z}^{(n)}$ as a $\boldsymbol{\theta}$-residual.

7.2 Maximal invariance of $\mathbf{F}^{(n)}_{\pm}$

Much attention always has been given in the literature to the invariance and equivariance properties of multivariate quantile and related concepts (see, e.g., Serfling (2010)). Empirical quantiles and ranks typically are expected to be equivariant and invariant, respectively, under the class $\mathcal{G}^{(n)}$ of "order-preserving transformations". Looking at it more closely, however, such a property is somewhat tautological, as order-preserving transformations are precisely those for which orderings (hence, the ranks) are invariant, hence the quantiles equivariant.

In dimension $d = 1$, the order-preserving transformations $\mathfrak{g}^{(n)} \in \mathcal{G}^{(n)}$ (see (7.5)) factorize as $\mathfrak{g}^{\otimes n}$, where the collection of marginal $\mathfrak{g}$'s acting on $\mathbb{R}$ does not depend on $\mathbf{Z}^{(n)}$ since the ordering itself is canonically defined irrespective of $\mathbf{Z}^{(n)}$. In the general case ($d \geq 2$), empirical (center-outward) orderings, characterized by $\mathbf{F}^{(n)}_{\pm}$ and $\mathbf{Q}^{(n)}_{\pm}$, are data-driven, hence depend on the observation $\mathbf{Z}^{(n)} = (\mathbf{Z}^{(n)}_1, \ldots, \mathbf{Z}^{(n)}_n)$. Unsurprisingly, the class of empirical-order-preserving transformations of $\mathbb{R}^{nd}$, of the form $\mathcal{G}^{(n)}_{\mathbf{Z}^{(n)}}$, also is data-driven; as a consequence, it is no longer a group.

Throughout, let us assume that the regular grid used for the definition of empirical distribution functions is fixed. When dependence on $\mathbf{z}^{(n)}$ is to be emphasized, we write $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, $\bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, $\mathbf{Q}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, $\mathcal{C}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, $\mathbb{C}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, etc. Denoting by $\mathbf{F}^{\mathrm{P}}_{\pm}$ and $\mathbf{Q}^{\mathrm{P}}_{\pm}$ the center-outward distribution and quantile functions, respectively, associated with $\mathrm{P} \in \mathcal{P}_d$, let $\mathcal{F}^d_{\pm} := \{\mathbf{F}^{\mathrm{P}}_{\pm} | \mathrm{P} \in \mathcal{P}_d\}$ and $\mathcal{Q}^d_{\pm} := \{\mathbf{Q}^{\mathrm{P}}_{\pm} | \mathrm{P} \in \mathcal{P}_d\}$.

For any $\mathbf{z}^{(n)} \in \mathbb{R}^{nd}$ at which $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ is uniquely defined (which happens Lebesgue-a.e.), consider the class of transformations (acting on $\mathbb{R}^{nd}$)

$$\mathcal{G}^{(n)}_{\mathbf{z}^{(n)}} := \big\{\mathfrak{g}^{\otimes n} | \, \mathfrak{g} = \mathbf{Q}_{\pm} \circ \bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}\big\} \tag{7.6}$$

with $\mathbf{Q}_{\pm}$ ranging over $\mathcal{Q}^d_{\pm}$, and $\bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ over the collection of homeomorphic gradients of convex functions interpolating the n-tuple $\big(\mathbf{z}^{(n)}_1, \mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}(\mathbf{z}^{(n)}_1)\big), \ldots, \big(\mathbf{z}^{(n)}_n, \mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}(\mathbf{z}^{(n)}_n)\big)$ (the existence of which is established in del Barrio et al. (2018)); in the sequel, we call this a *homeomorphic interpolation of* $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$.

The maximal invariance property (HW*) of of $\mathbf{F}^{(n)}_{\pm}$ and the related equivariance of the corresponding empirical center-outward quantile contours and regions then take the following form.

Proposition 7.1 (HW*) *(i) The class of transformations* $\mathcal{G}^{(n)}_{\mathbf{Z}^{(n)}}$ *is,* $\mathbf{Z}^{(n)}$*-a.s., a class of order-preserving transformations of* $\mathbb{R}^{nd}$*, with maximal invariant* $\mathbf{F}^{(n)}_{\pm;\mathbf{Z}^{(n)}}$*: namely, except for a set of* $\mathbf{z}^{(n)}$ *and* $\mathbf{y}^{(n)}$ *values of measure zero,* $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}} = \mathbf{F}^{(n)}_{\pm;\mathbf{y}^{(n)}}$ *(equivalently, the ranks and signs of* $\mathbf{z}^{(n)} = (\mathbf{z}^{(n)}_1, \ldots, \mathbf{z}^{(n)}_n)$ *and those of* $\mathbf{z}^{(n)} = (\mathbf{z}^{(n)}_1, \ldots, \mathbf{y}^{(n)}_n)$ *coincide) iff* $\mathbf{y}^{(n)} = \mathfrak{g}^{\otimes n}\mathbf{z}^{(n)}$ *for some* $\mathfrak{g}^{\otimes n}$ *in* $\mathcal{G}^{(n)}_{\mathbf{z}^{(n)}}$.
(ii) Each class $\mathcal{G}^{(n)}_{\mathbf{z}^{(n)}}$ *(except for a set of* $\mathbf{z}^{(n)}$ *values of measure zero) is a generating class for* $\mathcal{P}^{(n)}_d$*, that is, for any* f*,* $h \in \mathcal{F}^d$*, there exists* $\mathfrak{g}^{\otimes n}$ *in* $\mathcal{G}^{(n)}_{\mathbf{z}^{(n)}}$ *such that* $\mathfrak{g}^{\otimes n}\#\mathrm{P}^{(n)}_f = \mathrm{P}^{(n)}_h$.

Turning to the quantile contours and regions, we have the following equivariance counterparts to the invariance properties of Proposition 7.1.

Proposition 7.2 *The empirical quantile contours and regions* $\mathcal{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}$ *and* $\mathbb{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}$ *are* $\mathbf{Z}^{(n)}$*-a.s. equivariant under the class of transformations* $\mathcal{G}^{(n)}_{\mathbf{Z}^{(n)}}$ *defined in* (7.6)*: except for a a set of values of Lebesgue measure zero, for any* $\mathfrak{g}^{\otimes n} \in \mathcal{G}^{(n)}_{\mathbf{Z}^{(n)}}$ *with* $\mathfrak{g}$ *of the form* $\mathfrak{g} = \mathbf{Q}_{\pm} \circ \bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{Z}^{(n)}}$,

$$\mathcal{C}^{(n)}_{\pm;\mathfrak{g}^{\otimes n}\mathbf{Z}^{(n)}}(j/n_R) = \mathbf{Q}_{\pm} \circ \mathbf{F}^{(n)}_{\pm;\mathbf{Z}^{(n)}}\big(\mathcal{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}(j/n_R)\big) \tag{7.7}$$

and

$$\mathbb{C}^{(n)}_{\pm;\mathfrak{g}^{\otimes n}\mathbf{Z}^{(n)}}(j/n_R) = \mathbf{Q}_{\pm} \circ \mathbf{F}^{(n)}_{\pm;\mathbf{Z}^{(n)}}\big(\mathbb{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}(j/n_R)\big) \tag{7.8}$$

for all $j = 0, \ldots, n_R$.

Finally, for the the population concepts $\mathbf{F}_{\pm}$ and $\mathbf{Q}_{\pm}$, and the population quantile contours and regions, the of order-preserving transformations quite naturally are distribution-specific, with the following obvious invariance/equivariance properties.

Proposition 7.3 *Let* $\mathfrak{g} = \mathfrak{g}_{\mathrm{PQ}} := \mathbf{Q}_{\pm}^{\mathrm{Q}} \circ \mathbf{F}_{\pm}^{\mathrm{P}}$, $\mathrm{P}, \mathrm{Q} \in \mathcal{P}_d$. *Then,*

$$\mathbf{F}_{\pm}^{\mathfrak{g}\#\mathrm{P}} \circ \mathfrak{g} = \mathbf{F}_{\pm}^{\mathrm{P}}, \qquad \mathbf{Q}_{\pm}^{\mathfrak{g}\#\mathrm{P}} = \mathfrak{g} \circ \mathbf{Q}_{\pm}^{\mathrm{P}},$$
$$\mathfrak{g}\mathbb{C}^{\mathrm{P}}(q) = \mathbb{C}^{\mathrm{Q}}(q) \qquad \textit{and} \qquad \mathfrak{g}\mathcal{C}^{\mathrm{P}}(q) = \mathcal{C}^{\mathrm{Q}}(q), \qquad q \in (0,1).$$

8 Proofs

8.1 Proof of Proposition 4.1.

Parts (i) and (ii) of Proposition 4.1 are borrowed from Theorem 1.1 and the subsequent comments in Figalli (2018). Part (iii) follows from Equation (1.2) (same reference) and the fact that U_d has Lebesgue density $f_{\mathrm{U}}(\mathbf{u}) = \big[1/c_d\|\mathbf{u}\|^{d-1}\big]I[\mathbf{u} \in \mathbb{S}_d \setminus \{\mathbf{0}\}]$. □

8.2 Proof of Propositions 5.1 and 5.2 (Glivenko-Cantelli)

Let us start with some preliminary lemmas, all from McCann (1995). Throughout this section, μ and ν denote probability measures on $\mathbb{R}^d$, $\mathcal{P}(\mathbb{R}^d)$ the set of all probability distributions on $\mathbb{R}^d$, $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ the set of all probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$, and $\Gamma(\mu, \nu)$ the set of probability distributions in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with given marginals μ and ν in $\mathcal{P}(\mathbb{R}^d)$. A measure γ in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is said to have *cyclically monotone support* if there exists a closed Borel set S in $\mathbb{R}^d \times \mathbb{R}^d$ such that $\gamma(S) = 1$ and S is cyclically monotone.

Lemma 8.1 (McCann 1995, Corollary 14). *Let* μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$, *and suppose that one of those two measures vanishes on all sets of Hausdorff dimension* $d-1$. *Then, there exists one and only one measure* $\gamma \in \Gamma(\mu, \nu)$ *having cyclically monotone support.*

Lemma 8.2 (McCann 1995, Lemma 9). *Let* $\gamma^{(n)} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ *converge weakly as* $n \to \infty$ *to* $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. *Then,*

(i) if $\gamma^{(n)}$ *has cyclically monotone support for all* n, *so does* γ;

(ii) if $\gamma^{(n)} \in \Gamma(\mu^{(n)}, \nu^{(n)})$ *where* $\mu^{(n)}$ *and* $\nu^{(n)}$ *converge weakly, as* $n \to \infty$, *to* μ *and* ν, *respectively, then* $\gamma \in \Gamma(\mu, \nu)$.

Next, recall that the subdifferential $\partial\psi$ of a convex function $\psi : \mathbb{R}^d \to \mathbb{R}$ is the collection of pairs $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ such that $\psi(\mathbf{z}) \geq \psi(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle$, $\mathbf{z} \in \mathbb{R}^d$, that is, such that $\psi(\mathbf{z})$ lies entirely "above" the (supporting) hyperplane $\{\mathbf{z} : \mathbf{y}'(\mathbf{z} - \mathbf{x}) = 0\}$; $\mathbf{y}$ is called a *subgradient* of ψ at $\mathbf{x}$. A convex function being Lebesgue-a.e. differentiable, the subdifferential of a convex function ψ coincides Lebesgue-a.e. with the collection $\{(\mathbf{x}, \nabla\psi(\mathbf{x}))\}$.

Lemma 8.3 (McCann 1995, Proposition 10). *Suppose that* $\gamma \in \Gamma(\mu, \nu)$ *is supported on the subdifferential* $\partial\psi$ *of some convex function* ψ *on* $\mathbb{R}^d$ *(meaning that the support of* γ *is a subset of* $\partial\psi$*). Assume that* μ *vanishes on Borel sets of Hausdorff dimension* $d-1$. *Then,* $\nabla\psi$ *pushes* μ *forward to* ν, *that is,*

$$\gamma = (\text{ identity} \times \nabla\psi)\#\mu$$

where $(\text{ identity} \times \nabla\psi)\mathbf{x} := (\mathbf{x}, \nabla\psi(\mathbf{x}))$.

Finally, the following lemma by Rockafellar (1966) establishes a strong relation between cyclical monotonicity and convex functions (Rockafellar's statement actually holds for more general topological vector space).

Lemma 8.4 (Rockafellar 1966, Theorem 1). *The subdifferential $\partial\psi$ of a convex function ψ on $\mathbb{R}^d$ enjoys cyclical monotonicity. Conversely, any cyclically monotone set S of $\mathbb{R}^d \times \mathbb{R}^d$ is contained in the subdifferential $\partial\psi$ of some convex function ψ on $\mathbb{R}^d$.*

This implies the existence of a gradient of convex function running through any n-tuple of cyclically monotone couples $((\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)) \in \mathbb{R}^d \times \mathbb{R}^d$. We now turn to the proof of the Glivenko-Cantelli result (5.1) of Proposition 5.1.

Proof of Proposition 5.1. Let the n-tuple $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d$ and center-outward distribution function $\mathbf{F}_\pm$. Denote by $(\Omega, \mathcal{A}, \mathrm{P})$ the (unimportant) probability space underlying the observation of the sequence of $\mathbf{Z}_i^{(n)}$'s, $n \in \mathbb{N}$, by $\gamma^{(n)} = (\text{identity} \times \mathbf{F}_\pm^{(n)})\#\mu^{(n)}$ the empirical distribution of the couples $(\mathbf{Z}_i^{(n)}, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)}))$, with marginals $\mu^{(n)}$ and $U^{(n)}$, and by $\gamma = (\text{identity} \times \mathbf{F}_\pm)\#\mathrm{P}$ (with marginals P, U_d) the joint distribution of $(\mathbf{Z}, \mathbf{F}_\pm(\mathbf{Z}))$. Here, $\mu^{(n)}$, hence also $\gamma^{(n)}$ are random measures, with realizations $\mu_\omega^{(n)}$ and $\gamma_\omega^{(n)}$.

A sequence $\gamma_\omega^{(n)}$, $n \in \mathbb{N}$, is P-a.s. asymptotically tight since $\mu_\omega^{(n)}$ converges weakly to P with probability one, and $\mathrm{U}^{(n)}$ has uniformly bounded support. By Prohorov's theorem, subsequences $\gamma_\omega^{(n_k)}$ can be extracted that converge weakly (to some γ_ω^∞'s).

Those $\gamma_\omega^{(n_k)}$'s by construction have cyclically monotone supports, and their marginals $\mu_\omega^{(n_k)}$ and $\mathrm{U}^{(n_k)}$ converge weakly to P and U_d. Hence, by Lemma 8.2, all limiting γ_ω^∞'s have cyclically monotone supports, and marginals P and U_d, respectively.

In view of Lemma 8.1, there exists only one γ with cyclically monotone supports and marginals P and U_d. Hence, irrespective of the choice of the weakly converging subsequence $\gamma_\omega^{(n_k)}$, all limiting γ_ω^∞'s coincide with γ, which implies that the original sequence is converging weakly to γ. Moreover, that limit is the same for any ω in some $\Omega_1 \subseteq \Omega$ such that $\mathrm{P}(\Omega_1) = 1$.

Rockafellar's Theorem (Lemma 8.4) provides a convex function ψ the subgradient of which contains the support of γ. Lemma 8.3 and the definition of $\mathbf{F}_\pm$ concludes that

$$\gamma = (\text{identity} \times \nabla\psi)\#\mathrm{P} = (\text{identity} \times \mathbf{F}_\pm)\#\mathrm{P}.$$

Summing up, we have proved—mainly, by reorganizing elements of McCann's own proofs—the following result.

Lemma 8.5 *Let $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d$ and center-outward distribution function $\mathbf{F}_\pm$. Let $\mu^{(n)}$ be the corresponding empirical distribution, and $\mathbf{F}_\pm^{(n)}$ the corresponding empirical center-outward distribution function. As $n \to \infty$,*

$$\gamma^{(n)} := (\text{ identity} \times \mathbf{F}_\pm^{(n)})\#\mu^{(n)} \text{ converges weakly to } \gamma = (\text{identity} \times \mathbf{F}_\pm)\#\mathrm{P} \quad \mathrm{P}-a.s.$$

Interesting as it is, this result is only about almost sure *weak* convergence, which is not sufficient for a Glivenko-Cantelli result. To proceed further, let us turn to polar coordinates, and consider separately the Glivenko-Cantelli behaviour of $\|\mathbf{F}_\pm^{(n)}\|$ and that of the spherical coordinates of $\mathbf{F}_\pm^{(n)}/\|\mathbf{F}_\pm^{(n)}\|$.

Writing $\tau\mathbb{S}$ for the open ball $\tau\mathbb{S}_d$, consider the set of indicators (defined on $\mathbb{R}^d \times \mathbb{S}_d$)

$$\mathcal{F}_{\|\cdot\|} := \Big\{ f_\tau := I[\mathbf{F}_\pm^{-1}(\tau\mathbb{S}) \times \tau\mathbb{S}],\ \tau \in (0,1) \Big\},$$

and let us show that $\mathcal{F}_{\|\cdot\|}$ is a P-Glivenko-Cantelli class for any absolutely continuous P, that is,

$$\sup_{\tau\in[0,1)} \|\gamma^{(n)}(f_\tau) - \gamma(f_\tau)\| \longrightarrow_{n\to\infty} 0 \qquad \mathrm{P}-a.s.$$

where $\gamma^{(n)}(\cdot)$ and $\gamma(\cdot)$, as usual, denote expectations. It readily follows from Lemma 8.5 that

$$\left|\gamma^{(n)}(f_\tau) - \gamma(f_\tau)\right| = \left|\gamma^{(n)}(\mathbf{F}_\pm^{-1}(\tau\mathbb{S})\times\tau\mathbb{S}) - \gamma(\mathbf{F}_\pm^{-1}(\tau\mathbb{S})\times\tau\mathbb{S})\right|$$

converges to zero P-a.s. for any $\tau\in(0,1)$. To establish the uniformity over $\tau\in(0,1)$ of this convergence, consider, for $\epsilon>0$,

$$t_1,\, t_2,\, \ldots,\, t_{\lceil 1/\epsilon\rceil} := 0,\, \epsilon,\, 2\epsilon,\, \ldots,\, 1$$

such that $t_j - t_{j-1} = \mathrm{U}_d\big(t_j\mathbb{S}\setminus t_{j-1}\mathbb{S}\big)\le\epsilon$, and the brackets $(f_{t_{j-1}}, f_{t_j}]$. Those brackets have P-, hence γ-size at most ϵ. Their total number can be chosen less than $(1/\epsilon)+1$. That number is finite for any $\epsilon>0$, which entails uniformity:

$$\sup_\tau\left|\gamma^{(n)}(\mathbf{F}_\pm^{-1}(\tau\mathbb{S})\times\tau\mathbb{S}) - \gamma(\mathbf{F}_\pm^{-1}(\tau\mathbb{S})\times\tau\mathbb{S})\right| \longrightarrow 0 \quad \text{P-a.s., } n\to\infty$$

hence also

$$\max_{1\le i\le n}\left|\gamma^{(n)}\Big(\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big) - \gamma\Big(\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big)\right| \qquad (8.9)$$
$$\longrightarrow 0 \quad \text{P-a.s., } n\to\infty.$$

Now, by definition,

$$\gamma\Big(\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big) = \|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|, \qquad (8.10)$$

and

$$\gamma^{(n)}\Big(\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big) \qquad (8.11)$$
$$= \gamma^{(n)}\Big(\mathbf{F}_\pm^{(n)-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\bigcap\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\times\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big).$$

Together, (8.9) and (8.11) entail

$$\max_{1\le i\le n}\gamma^{(n)}\Big(\mathbf{F}_\pm^{(n)-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\setminus\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\Big) = o(1) \quad \text{P-a.s., } n\to\infty$$

hence

$$\max_{1\le i\le n}\left|\gamma^{(n)}\Big(\mathbf{F}_\pm^{(n)-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\bigcap\mathbf{F}_\pm^{-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\times\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big)\right. \qquad (8.12)$$
$$\left.-\gamma^{(n)}\Big(\mathbf{F}_\pm^{(n)-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big)\right| = o(1) \quad \text{P-a.s., } n\to\infty$$

But

$$\gamma^{(n)}\Big(\mathbf{F}_\pm^{(n)-1}(\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S})\times\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\mathbb{S}\Big) = \|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|; \qquad (8.13)$$

the claim thus follows from piecing together (8.9), (8.10), (8.12) and (8.13).

A similar reasoning based on Lemma 8.5 and a bracketing argument can be invoked for the parallel, meridian and hypermeridian coordinates of $\mathbf{F}_\pm/\|\mathbf{F}_\pm\|$ and $\mathbf{F}_\pm^{(n)}/\|\mathbf{F}_\pm^{(n)}\|$. Recall that a point $\mathbf{u}$ on the unit sphere $\mathcal{S}_{d-1}$ is represented, in hyperspherical coordinates, by $d-2$ angles $\varsigma_1(\mathbf{u}),\ldots,\varsigma_{d-2}(\mathbf{u})$ ranging over $[0,\pi]$ and one angle $\varsigma_{d-1}(\mathbf{u})$ ranging over $[0,2\pi)$. Let $\mathbf{e}_1,\ldots,\mathbf{e}_d$ denote an arbitrary orthonormal basis of $\mathbb{R}^d$. That basis characterizes a hyperspherical system as follows: $\varsigma_j(\mathbf{u})$ is a polar angle measured from the zenith direction $\mathbf{e}_j$, $j=1,\ldots,d-2$, while $\varsigma_{d-1}(\mathbf{u})$ is an azimuth measured, in the hyperplane spanned by $\mathbf{e}_{d-1}$ and $\mathbf{e}_d$, from the azimuth reference direction $\mathbf{e}_{d-1}$. The correspondence between the Euclidean coordinate system based on $\mathbf{e}_1,\ldots,\mathbf{e}_d$ and the hyperspherical one is

$$\begin{aligned}
\mathbf{u}'\mathbf{e}_1 &= \cos(\varsigma_1)\\
\mathbf{u}'\mathbf{e}_2 &= \sin(\varsigma_1)\cos(\varsigma_2)\\
\ldots &= \ldots\\
\mathbf{u}'\mathbf{e}_{d-1} &= \sin(\varsigma_1)\sin(\varsigma_2)\ldots\sin(\varsigma_{d-3})\sin(\varsigma_{d-2})\cos(\varsigma_{d-1})\\
\mathbf{u}'\mathbf{e}_d &= \sin(\varsigma_1)\sin(\varsigma_2)\ldots\sin(\varsigma_{d-3})\sin(\varsigma_{d-2})\sin(\varsigma_{d-1}).
\end{aligned}$$

A Glivenko-Cantelli result for $\varsigma_j^{(n)}$, namely,

$$\max_{1\le i\le n}\left|\varsigma_{i;j}^{(n)}-\varsigma_{i;j}\right|\longrightarrow 0\quad\text{a.s. as } n\to\infty$$

where $\varsigma_{i;j}^{(n)}$ and $\varsigma_{i;j}$ stand for the polar coordinates

$$\varsigma_j\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})/\|\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\|\Big)\quad\text{and}\quad\varsigma_j\Big(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})/\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|\Big),$$

respectively, then holds for all $1\le j\le d-2$. This follows along the same lines as for $\|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)})\|$, with nested hyperspherical caps (with axes $\mathbf{e}_1,\ldots,\mathbf{e}_{d-2}$, respectively) replacing the nested hyperspheres. For $j=d-1$, the nested caps are replaced by nested hyperspherical lunes, i.e. hyperspherical domains comprised between two hyperplanes intersecting along $\mathbf{e}_{d-2}$, one of them containing $\mathbf{e}_{d-1}$, the other one forming with the latter a dyhedral angle ς_{d-1}. □

Proof of Proposition 5.2. Proposition 5.2 is an immediate consequence of Proposition 5.1, the almost-sure continuous mapping theorem, and the fact (Section 2.4 in Chernozhukov et al. (2017)) that, for spherical $\mathbf{Z}$, $\mathbf{F}_{\text{ell}}$ coincides with $\mathbf{F}_\pm$; details are left to the reader. □

8.3 Proof of Proposition 6.1 (Distribution-freeness)

Starting with Part (i), let $\mathbf{S}^{(n)} := (S_{1;1}^{(n)},\ldots,S_{1;n}^{(n)},S_{2;1}^{(n)},\ldots,S_{2;n}^{(n)},S_{d;1}^{(n)},\ldots,S_{d;n}^{(n)},S_{12}^{(n)},\ldots,S_{1d}^{(n)})$, with $S_{j;k}^{(n)} := \sum_{i=1}^n (Z_{ij}^{(n)})^k$ for $j=1,\ldots,d$ and $k=1,\ldots,n$, and $S_{1j}^{(n)} := \sum_{i=1}^n Z_{i1}^{(n)}Z_{ij}^{(n)}$ for $j=1,\ldots,d$. It is easy to see, along the lines of in Example 2.4.1 of Lehmann and Romano (2005), that $\mathbf{Z}_{(\,.\,)}^{(n)}$ and $\mathbf{S}^{(n)}$ both induce the same sub-σ-field of the observation space $\mathbb{R}^{nd}$. Hence, $\mathbf{Z}_{(\,.\,)}^{(n)}$ is sufficient and complete iff $\mathbf{S}^{(n)}$ is. It follows from the Fisher factorization criterion that $\mathbf{Z}_{(\,.\,)}^{(n)}$ is sufficient for the model of n i.i.d. observations with distribution $\mathrm{P}\in\mathcal{P}^d$. That model contains the parametric submodel of n i.i.d. observations with exponential distribution and complete minimal sufficient statistic $\mathbf{S}^{(n)}$. Minimal sufficiency and completeness therefore carry over to $\mathbf{Z}_{(\,.\,)}^{(n)}$ and the full model.

Turning to Part (ii), assume, for simplicity, that $n_0 = 0$ or 1 (else, replace the origin with n_0 undistinguishable copies, and $n!$ with the number $n!/n_0!$ of permutations with repetitions). That the distribution of $\big(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}\big)$ is uniform over the $n!$ permutations of the augmented grid is a consequence of the optimal pairing between the observations and the augmented grid actually is an optimal pairing between the elements of the order statistic and the augmented grid. Conditionally on the order statistic $\mathbf{Z}_{(\,.\,)}^{(n)}$, the observations are uniformly distributed over the $n!$ permutations of $\mathbf{Z}_{(\,.\,)}^{(n)}$, so that the conditional distribution of the $\mathbf{F}^{(n)}(\mathbf{Z}_i^{(n)})$'s is (almost surely) uniform over the $n!$ permutations of the grid. Since that conditional distribution does not depend on the conditioning variable, it is also unconditional.

Point (iii) finally is an immediate consequence of point (ii) and the classical Basu Theorem (Basu 1955). □

8.4 Proof of Propositions 7.1 and 7.2 (Invariance/equivariance)

Proof of Proposition 7.1. (i) Let $\mathbf{z}^{(n)} \in \mathbb{R}^{nd}$ and $\mathbf{y}^{(n)} \in \mathbb{R}^{nd}$ be such that $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ and $\mathbf{F}^{(n)}_{\pm;\mathbf{y}^{(n)}}$ are well defined (which happens Lebesgue-a.e.); then, $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}(\mathbf{z}^{(n)}) = \mathbf{F}^{(n)}_{\pm;\mathbf{y}^{(n)}}(\mathbf{y}^{(n)})$ iff $\mathbf{y}^{(n)} = \mathfrak{g}^{\otimes n}\mathbf{z}^{(n)}$ with $\mathfrak{g} = \bar{\mathbf{Q}}^{(n)}_{\pm;\mathbf{y}^{(n)}} \circ \bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, where $\bar{\mathbf{Q}}^{(n)}_{\pm;\mathbf{y}^{(n)}}$ and $\bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ are arbitrary homeomorphic interpolations of $\mathbf{Q}^{(n)}_{\pm;\mathbf{y}^{(n)}}$ and $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$, respectively, so that $\mathfrak{g}^{\otimes n}$ is an element of the class $\mathcal{G}^{(n)}_{\mathbf{z}^{(n)}}$.

(ii) Choose $\mathfrak{g} = \mathbf{Q}_\pm \circ \bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ where $\bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ is an arbitrary homeomorphic interpolations of $\mathbf{F}^{(n)}_{\pm;\mathbf{z}^{(n)}}$ and $\mathbf{Q}_\pm$ is the unique gradient of convex function pushing $\bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{z}^{(n)}}\#\mathrm{P}_f$ (an element of $\mathcal{P}$) forward to P_h. □

Proof of Proposition 7.2. Denoting by $\mathbf{u}_i^{(n)}$, $i = 1, \ldots, n$ the n gridpoints, let

$$\mathcal{S}^{(n)}(j/n_R) := \big\{\mathbf{u}_i^{(n)} \,|\, \|\mathbf{u}_i^{(n)}\| = j/(n_R+1)\big\}, \qquad j = 1, \ldots, n_R.$$

Put $\mathbf{F}_\pm := (\mathbf{Q}_\pm)^{-1}$ and write $\mathbf{Y}_i^{(n)} := \mathbf{Q}_\pm \circ \bar{\mathbf{F}}^{(n)}_{\pm;\mathbf{Z}^{(n)}}\mathbf{Z}_i^{(n)}$, $i = 1, \ldots, n$. Since $\mathbf{F}_\pm$ and $\mathbf{F}^{(n)}_{\pm;\mathbf{Y}^{(n)}}$ coincide on the $\mathbf{Y}_i^{(n)}$'s, (7.7) holds iff

$$\mathbf{F}_{\pm;\mathbf{Y}^{(n)}}\mathcal{C}^{(n)}_{\pm;\mathbf{Y}^{(n)}}\big(j/n_R\big) = \mathbf{F}^{(n)}_{\pm;\mathbf{Z}^{(n)}}\big(\mathcal{C}^{(n)}_{\pm;\mathbf{Z}^{(n)}}(j/n_R)\big), \quad j = 1, \ldots, n_R. \tag{8.14}$$

This latter equality holds true, as both sides in (8.14) by definition reduce to $\mathcal{S}^{(n)}(j/n_R)$. The result follows, and readily implies (7.8). □

9 Conclusions

The concepts of distribution and quantile functions, ranks and signs, are well understood, essentially, in dimension one and in elliptical families, where they enjoy distribution-freeness and allow (Hallin and Werker 2003) for the construction of semiparametrically efficient inference procedures (tests and R-estimation) in models involving the unspecified density of some unobserved residual noise. A measure transportation-based characterization of a center-outward form of those univariate concepts readily extends to the d-dimensional case. We show that the resulting new concepts of distribution and quantile functions, ranks and signs enjoy in $\mathbb{R}^d$ the properties that make their traditional versions successful inferential tools in the univariate case.

In principle, our concepts open the door to a new theory of empirical processes, calling for further results such as Donsker and iterated logarithm theorems, or Bahadur representations. They also pave the way to a solution of the long-standing open problem of rank-based inference in multivariate analysis in the absence of "any" distributional assumptions, offering a combination of strict distribution-freeness and semiparametric efficiency none of the previous concepts of multivariate ranks and signs can offer.

Many questions remain open, though, until those objectives can be attained.

(i) Several issues remain to be studied about the concepts themselves: how in finite samples should we choose the factorization into $n_R n_S + n_0$? should we combine several of them? should we consider cross-validation? how? how should we smooth the discretely defined quantile contours? what happens if we drop the assumption of nonvanishing densities?

(ii) How exactly should we construct efficient rank tests in specific problems? Proposition 5.2 suggests replacing, in the many test statistics derived, under elliptic symmetry, by Hallin, Paindaveine and Verdebout (see the references below), the Mahalanobis (elliptical) ranks and signs with the center-outward ones. Can we similarly construct one-step R-estimators (a problem which, for $d \geq 2$, so far is solved under elliptical symmetry only)? This would result in a fairly complete toolkit of distribution-free (hence "universally valid") semiparametrically efficient-at-elliptical-densities rank-based inference procedures for multivariate analysis and multivariate time series problems.

(iii) Can goodness-of-fit tests be based, e.g. on Kolmogorov-Smirnov or Cramér-von Mises distances between center-outward distribution functions?

(iv) Turning to quantiles, what are the properties of $\mathbf{Q}_{\pm}^{(n)}(0)$ as a multivariate median? can we construct multivariate median tests? can we, on the model of Carlier et al. (2016) or Hallin et al. (2010, 2015), perform multiple-output quantile regression (reconstruction of conditional center-outward quantile contours as a function of covariates)? construct multivariate growthcharts (as in McKeague et al. (2011))? How?

(v) Center-outward quantile contours are obvious candidates as multivariate value-at-risk concepts, playing a central role in risk management; in that context, still in dimension $d = 1$, the primitives of ordinary distribution or quantile functions enter the definitions of a number of relevant quantities such as Lorenz curves, average values at risk, or expected shortfall, see Gushchin and Borzykh (2017). The potential function Ψ characterizing the underlying $\mathbf{F}_{\pm}$ and its Legendre transform are natural multivariate extensions of those primitives, and likely to provide useful generalizations of those concepts.

(vi) Finally, what happens in high dimension? in functional spaces? on spheres (directional data)? on other Riemannian manifolds?

References

[1] Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhya* **15**, 377–380.

[2] Belloni, A. and Winkler, R.L. (2011). On multivariate quantiles under partial orders. *Annals of Statistics* **39**, 1125–1179.

[3] Bickel, P.J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2, *The Annals of Mathematical Statistics* **36**, 160-173.

[4] Bickel, P.J. , Klaassen, Ritov, and Wellner, J. (1993). *Efficient and Adaptive Inference in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

[5] Biswas, M., Mukhopadhyay, M. and Ghosh, A.K. (2014). A distribution-free two-sample run test applicable to high-dimensional data, *Biometrika* **101**, 913–926.

[6] Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs, *Comptes Rendus de l'Académie des Sciences de Paris Série I Mathémathique* **305** **19**, 805–808.

[7] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions, *Communications in Pure and Applied Mathematics* **44**, 375–417.

[8] Carlier, Chernozhukov, V. and Galichon, A. (2016). Vector quantile regression, *Annals of Statistics* **44**, 1165–1192.

[9] Chakraborty and Chaudhuri (1996). On a transformation-retransformation technique for constructing an affine equivariant multivariate median, *Proceedings of the American Mathematical Society* **124**, 2539–2547.

[10] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths, *Annals of Statistics* **42**, 1203–1231.

[11] Chakraborty, A. and Chaudhuri, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks, *Annals of Statistics* **45**, 771–799.

[12] Chakraborty, B. and Chaudhuri, P. (1996). On a transformation and retransformation technique for constructing affine equivariant multivariate median, *Proceedings of the American Mathematical Society* **124**, 2539–2547.

[13] Chakraborty, B. and Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location, *Journal of the Royal Statistical Society Series B* **60**, 145–157.

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

[15] Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension : inference based on the geometry of the data cloud, *Journal of the American Statistical Association* **88**, 1363–1370.

[16] Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *The Annals of Mathematical Statistics* **29**, 972–994.

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

[18] Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance, *Journal of the American Statistical Association* **92**, 1581–1590.

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

[20] Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration, in W. Chen, A. Srivastava and G. Travaglini, Eds, *A Panorama of Discrepancy Theory*, Lecture Notes in Mathematics 2107, 539–620, Springer, New York.

[21] Ekeland, I., Galichon, A. and Henrt, M. (2012). Comonotonic measures of multivariate risks, *Mathematical Finance* **22**, 109–132.

[22] Figalli, A. (2017). On the continuity of center-outward distribution and quantile functions, arXiv:1805.04946.

[23] Galichon, A. (2016). *Optimal Transport Methods in Economics*, Princeton University Press, Princeton, N.J.

[24] Genest, Chr. and Rivest, P.L. (2001). On the multivariate probability integral transformation, *Statistics & Probability Letters* **53**, 391–399.

[25] Gushchin, A.A. and Borzykh, D.A. (2017). Integrated quantile functions: properties and applications, *Modern Stochastics, Theory and Applications* **4**, 285–314.

[26] Hallin, M. (1994). On the Pitman nonadmissibility of correlogram-based time series methods, *Journal of Time Series Analysis* **16**, 607–612.

[27] Hallin, M., Ingenbleek, J-Fr. and Puri, M.L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *Journal of Multivariate Analysis* **30**, 34–71.

[28] Hallin, M. and La Vecchia, D. (2018). A simple R-estimation method for semiparametric duration models, *Journal of Econometrics*, to appear.

[29] Hallin, M., Lu, Z., Paindaveine, D. and Šiman, M. (2015). Local bilinear multiple-output quantile regression, *Bernoulli* **21**, 1435–1466.

[30] Hallin, M. and Mehta, Ch. (2015). R-Estimation for asymmetric independent component analysis. *Journal of the American Statistical Association* **110**, 218–232.

[31] Hallin, M., Oja, H. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape, *Annals of Statistics* **34**, 2757–2789.

[32] Hallin, M. and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics* **30**, 1103–1133.

[33] Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8**, 787–815.

[34] Hallin, M. and Paindaveine, D. (2002c). Multivariate signed ranks : Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1 Norm and Related Procedures*, Y. Dodge, Ed., Birkhäuser, 271–282.

[35] Hallin, M. and Paindaveine, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Annals of Statistics* **32**, 2642–2678.

[36] Hallin, M. and Paindaveine, D. (2004b). Multivariate signed rank tests in vector autoregressive order identification. *Statistical Science* 19, 697–711.

[37] Hallin, M. and Paindaveine, D. (2005a). Affine-invariant aligned rank tests for the multivariate general linear model with ARMA errors, *Journal of Multivariate Analysis* **93**, 122–163.

[38] Hallin, M. and Paindaveine, D. (2005b). Asymptotic linearity of serial and nonserial multivariate signed rank statistics, *Journal of Statistical Planning and Inference* **136**, 1–32.

[39] Hallin, M. and Paindaveine, D. (2006a). Semiparametrically efficient rank-based inference for shape: I Optimal rank-based tests for sphericity, *Annals of Statistics* **34**, 2707–2756.

[40] Hallin, M. and Paindaveine, D. (2006b). Parametric and semiparametric inference for shape: the role of the scale functional, *Statistics & Decisions* **24**, 1001–1023.

[41] Hallin, M. and Paindaveine, D. (2008a). Chernoff-Savage and Hodges-Lehmann results for Wilks test of multivariate independence. In M. Silvapulle, Ed., *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, I.M.S. Lecture Notes-Monograph Series, 184–196.

[42] Hallin, M. and Paindaveine, D. (2008b). Optimal rank-based tests for homogeneity of scatter, *Annals of Statistics* **36**, 1261-1298.

[43] Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L^1 optimization to halfspace depth" (with discussion), *Annals of Statistics* **38**, 635–669.

[44] Hallin, M., Paindaveine, D. and Verdebout, Th. (2010). Optimal rank-based testing for principal components, *Annals of Statistics* **38**, 3245–3299.

[45] Hallin, M., Paindaveine, D. and Verdebout, Th. (2013). Optimal rank-based tests for common principal components, *Bernoulli* **19**, 2524–2556.

[46] Hallin, M., Paindaveine, D. and Verdebout, Th. (2014). Efficient R-estimation of principal and common principal components, *Journal of the American Statistical Association* **109**, 1071–1083.

[47] Hallin, M. and Tribel, O. (2000). The efficiency of some nonparametric competitors to correlogram-based methods. In F.T. Bruss and L. Le Cam, Eds, *Game Theory, Optimal Stopping, Probability, and Statistics, Papers in honor of T.S. Ferguson on the occasion of his 70th birthday*, I.M.S. Lecture Notes-Monograph Series, 249–262.

[48] Hallin, M. and Werker, B.J.M. (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 137–165.

[49] Hamel, A. and Kostner, D. (2016). Cone distribution functions and quantiles for multivariate random variables, arXiv:1611.06353.

[50] He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets, *Annals of Statistics* **25**, 495–504.

[51] Hodges, J.L. (1955). A bivariate sign test, *The Annals of Mathematical Statistics* **26**, 523-527.

[52] Hodges, J.L. and Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t test, *The Annals of Mathematical Statistics* **27**, 324–335.

[53] Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models, *Annals of Statistics* **39**, 2448–2476.

[54] Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, MA.

[55] Koltchinskii, V. (1997). M-Estimation, convexity and quantiles, *Annals of Statistics* **25**, 435–477.

[56] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd Edition, Springer, New York.

[57] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, Springer, New York.

[58] Liu, R. Y. (1992). Data depth and multivariate rank tests, in Y. Dodge, Ed., *L^1 Statistics and Related Methods*, 279–294, North-Holland, Amsterdam.

[59] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *Journal of the American Statistical Association* **88**, 257–260.

[60] Marden, J. (1999). Multivariate rank tests., in S. Ghosh, Ed., *Design of Experiments and Survey Sampling*, M. Dekker, New York, 401–432.

[61] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps, *Duke Mathematical Journal* **80**, 309–324.

[62] McKeague, I.W., López-Pintado, S., Hallin, M. and Šiman, M. (2011). Analyzing growth trajectories. *Journal of Developmental Origins of Health and Disease* **2**, 322-329.

[63] Monge, G. (1781) *Mémoire sur la Théorie des Déblais et des Remblais*, *Mémoires de l'Académie Royale des Sciences* **1781**, 666–704.

[64] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics* **5**, 201-213.

[65] Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Annals of Statistics* **25**, 542–552.

[66] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, SIAM Society for Industrial and Applied Mathematics, Philadelphia, PA.

[67] Nordhausen, K., Oja, H., and Paindaveine, D. (2009). Signed-rank tests for location in the symmetric independent component model. *Journal of Multivariate Analysis* **100**, 821–834.

[68] Nordhausen, K., Oja, H., and Tyler, D.E. (2006). On the efficiency of invariant multivariate sign and rank tests, in Liski, E.P., Isotalo, J., Niemelä, J., Puntanen, S., and Styan, G.P.H., Eds, *Festschrift for Tarmo Pukkila on his 60th Birthday*, 217-231, University of Tampere, Tampere.

[69] Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scandinavian Journal of Statistics* **26**, 319–343.

[70] Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks*. Springer, New York.

[71] Oja, H., and Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes, *Journal of Statistical Planning and Inference* **135**, 300–323.

[72] Oja, H., and Randles, R.H. (2004). Multivariate nonparametric tests. *Statistical Science* **19**, 598–605.

[73] Paindaveine, D. (2004). A unified and elementary proof of serial and nonserial, univariate and multivariate, Chernoff- Savage results. *Statistical Methodology* **1**, 81–91.

[74] Paindaveine, D. (2006). A Chernoff-Savage result for shape. On the non-admissibility of pseudo-Gaussian methods. *Journal of Multivariate Analysis* **97**, 2206–2220.

[75] Paindaveine, D. (2007). A canonical definition of shape. *Statistics & Probability Letters* **78**, 2240–2247

[76] Panaretos, V. and Zemel, Y. (2016). Amplitude and phase variation of point processes. *Annals of Statistics* **44**, 771–812.

[77] Panaretos, V. and Zemel, Y. (2018). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli*, to appear.

[78] Puri, M.L. and Sen, P.K. (1966). On a class of multivariate multisample rank-order tests (1966). *Sankhya, Series A* **28**, 353–376.

[79] Puri, M.L. and Sen, P.K. (1967). On the theory of rank order tests for location in the multivariate one-sample problem. *The Annals of Mathematical Statistics* **38**, 1216–1228.

[80] Puri, M.L. and Sen, P.K. (1969). A class of rank order tests for a general linear hypothesis. *The Annals of Mathematical Statistics* **40**, 1325–1343.

[81] Puri, M.L. and Sen, P.K. (1970). Asymptotic theory of likelihood ratio and rank order tests in some multivariate linear models. *The Annals of Mathematical Statistics* **41**, 87–100.

[82] Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis.* John Wiley & Sons, New York.

[83] Rachev, S.T. and Rüschendorf, L. (1998a, b). *Mass Transportation Problems*, Volume I and II, Springer, New York.

[84] Randles, R.H. (1989). A distribution-free multivariate sign test based on interdirections, *Journal of the American Statistical Association* **84**, 1045–1050.

[85] Randles, R.H. (2000). A simpler, affine equivariant multivariate, distribution-free sign test. *Journal ofthe American Statistical Association* **95**, 1263–1268.

[86] Rockafellar, R.T. (1966). Characterization of the subdifferentials of convex functions, *Pacific Journal of Mathematics* **17**, 497–510.

[87] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

[88] Segers, J., van den Akker, R., and Werker, B.J.M. (2014). Semiparametric Gaussian copula models: geometry and efficient rank-based estimation, *Annals of Statistics* **42**, 1911–1940.

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

[90] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardazation, *Journal of Nonparametric Statistics* **22**, 915–936.

[91] Serfling (2012). Depth, in A.-H. El-Shaarawi and W. Piegorsch, Eds., *Encyclopedia of Environmetrics*, 2nd Edition, Wiley, 636–641.

[92] Um, Y. and Randles, R.H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statistica Sinica* **8**, 801–812.

[93] Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI.

[94] Villani, C. (2009). *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Heidelberg.

[95] Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin* **1**, 80–83.

[96] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function, *Annals of Statistics* **28**, 461–482.

[97] Zuo, Y. (2018). On general notions of depth for regression, arXiv:1805.02046v1.

[98] Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests, *Annals of Statistics* **34**, 2879–2896.

Marc Hallin
E.C.A.R.E.S., Université Libre de Bruxelles CP114/04
50 avenue F.D. Roosevelt
B-1050 Bruxelles
Belgium

mhallin@ulb.ac.be