



Measure Transportation and
Statistical Decision Theory

Marc Hallin
ECARES, Université libre de Bruxelles

January 2021

ECARES working paper 2021-04

Measure Transportation and Statistical Decision Theory

Marc Hallin

ECARES and Department of Mathematics

Université libre de Bruxelles, Brussels, Belgium, B-1050

email: mhallin@ulb.ac.be

February 20, 2021

Abstract

Unlike the real line, the real space, in dimension $d \geq 2$, is not canonically ordered. As a consequence, extending to a multivariate context fundamental univariate statistical tools such as quantiles, signs, and ranks is anything but obvious. Tentative definitions have been proposed in the literature but do not enjoy the basic properties (e.g., distribution-freeness of ranks, their independence with respect to the order statistic, their independence with respect to signs, etc.) they are expected to satisfy. Based on measure transportation ideas, new concepts of distribution and quantile functions, ranks, and signs have been proposed recently that, unlike previous attempts, do satisfy these properties. These ranks, signs, and quantiles have been used, quite successfully, in several inference problems and have triggered, in a short span of time, a number of applications: fully distribution-free testing for multiple-output regression, MANOVA, and VAR models, R-estimation for VARMA parameters, distribution-free testing for vector independence, multiple-output quantile regression, nonlinear independent component analysis, etc.

Contents

1	INTRODUCTION	2
2	MEASURE TRANSPORTATION IN A NUTSHELL	3
3	DISTRIBUTION AND QUANTILE FUNCTIONS, RANKS, AND SIGNS	4
3.1	Ranks in dimension $d = 1$	5
3.2	Ranks in dimension $d > 1$: a survey	6
3.3	Center-outward distribution and quantile functions, ranks, and signs	7
3.3.1	Center-outward distribution and quantile functions, ranks and signs, $d = 1$	8
3.3.2	Monge-Kantorovich depth, vector quantiles, and vector ranks, $d \geq 1$. . .	8
3.3.3	Center-outward distribution and quantile functions, ranks and signs, $d \geq 1$	9
3.4	Unit cube or unit ball?	14

4	APPLICATIONS IN STATISTICAL INFERENCE	15
4.1	Center-outward rank tests and R-estimators	15
4.1.1	Multiple-output regression models and MANOVA	16
4.1.2	VAR and VARMA models	17
4.1.3	Vector independence	18
4.1.4	Goodness-of-fit and symmetry	21
4.2	Multiple-output quantile regression	21
4.3	Nonlinear Principal Components (PC) and Independent Components (IC)	23
5	CONCLUSIONS AND SOME DIRECTIONS FOR FUTURE RESEARCH	24

1 INTRODUCTION

Unlike the real line, the real space, in dimension two and higher, is not canonically ordered. As a consequence, a number of fundamental univariate probabilistic concepts and essential statistical tools—distribution and quantile functions, signs, ranks, values-at-risk, expected shortfalls, survival probabilities, risk dominance, concentration indices, . . .—all strongly related to the canonical ordering of the real line—do not canonically extend to a multivariate setting.

This problem is not new, and many attempts have been made to provide solutions. In particular, several proposals have been made for multivariate concepts of ranks and signs, hence distribution and quantile functions. In a sense, the theory of copulas and the concept of statistical depth were motivated by the same need of a meaningful ordering of \mathbb{R}^d . A new approach, based on measure transportation ideas, has been taken recently (Chernozhukov et al. (2017), Hallin et al. (2021a)), yielding notions of *center-outward* distribution and quantile functions, ranks, and signs in \mathbb{R}^d for arbitrary d . Contrary to previous proposals, these measure-transportation-based notions enjoy all the properties that make their traditional univariate counterparts successful tools for statistical inference.

In the past decades, measure transportation has emerged as one of the most active and vigorous subjects of pure and applied mathematics, with successful applications in a number of fields, ranging from fluid dynamics to economics, operations research, probability, and machine learning. Statistics was somewhat slow to join that tendency. Statistical applications of Wasserstein distances came first. The Wasserstein distance indeed metricizes convergence in distribution, hence naturally comes into the picture as a powerful tool in the derivation of asymptotic results but also as an obvious instrument for goodness-of-fit testing. These statistical aspects of Wasserstein distances are carefully covered in a recent *Annual Reviews* (Panaretos and Zemel, 2019) and will not be considered here. Nor do we consider risk-analysis-, microeconomics, and tail-behavior-related papers such as Beirlant et al. (2020), Chernozhukov et al. (2020), or de Valk and Segers (2019). Rather, we concentrate on more recent developments where measure transportation is used as a way to define an ordering of \mathbb{R}^d leading to concepts of distribution and quantile functions, ranks, and signs, with applications in distribution-free

rank-based inference.

2 MEASURE TRANSPORTATION IN A NUTSHELL

Measure transportation goes back to Gaspard Monge (1746-1818) and his 1781 *Mémoire sur la Théorie des Déblais et des Remblais*.¹ Starting from a very practical problem—*How should one best move a given pile of sand (clearing) to fill up a given hole of the same total volume (filling)?*—Monge actually initiated a profound mathematical theory anticipating various areas of differential geometry, linear programming, nonlinear partial differential equations, fluid mechanics, and probability.

In mathematical notation and probabilistic form, the simplest formulation of Monge’s problem is as follows. Let P_1 and P_2 denote two probability measures over (for simplicity) $(\mathbb{R}^d, \mathcal{B}^d)$ and let $L : \mathbb{R}^{2d} \rightarrow [0, \infty]$ be a Borel-measurable loss function such that $L(\mathbf{x}_1, \mathbf{x}_2)$ represents the cost of transporting \mathbf{x}_1 to \mathbf{x}_2 . The objective is to find a measurable (transport) map $T_{P_1;P_2} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ solving the minimization problem

$$\inf_T \int_{\mathbb{R}^d} L(\mathbf{x}, T(\mathbf{x})) dP_1 \quad \text{subject to} \quad T\#P_1 = P_2 \quad (1)$$

where T ranges over the set of measurable map from \mathbb{R}^d to \mathbb{R}^d and $T\#P_1$ is the so-called *push forward of P_1 by T* .² For simplicity, and with a slight abuse of language, we will say that T is *mapping P_1 to P_2* . A map $T_{P_1;P_2}$ achieving the infimum in (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 $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$; problem (1) then takes the form

$$\inf_T \int_{\mathbb{R}^d} L(\|\mathbf{x} - T(\mathbf{x})\|_2^2) dP_1 \quad \text{subject to} \quad T\#P_1 = P_2 \quad (2)$$

of a minimization, over all measurable mappings (transports) T from \mathbb{R}^d to \mathbb{R}^d , of the expected squared distance traveled from the (random) original location $\mathbf{X} \sim P_1$ to the planned destination $T(\mathbf{X}) \sim P_2$.

Monge’s problem may look straightforward but it is not. The transportation cost to be minimized in (1) and (2) indeed is nonlinear in T and the set $\{T | T\#P_1 = P_2\}$ of admissible transports over which this cost is to be minimized is not convex. 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 1942 and the pathbreaking approach by Leonid Kantorovich (who in 1942 was unaware of the relation between his work and Monge’s). The problem Kantorovich was dealing with is the minimization

$$\inf_{\gamma} \int_{\mathbb{R}^d \times \mathbb{R}^d} L(\mathbf{x}, \mathbf{y}) d\gamma(\mathbf{x}, \mathbf{y}) \quad \text{subject to} \quad \gamma \in \Gamma(P_1, P_2) \quad (3)$$

¹ *Memoir on the Theory of Clearings and Fillings*.

² In statistics, a more classical but heavier notation for $T\#P_1$ would be $P_1^{T\mathbf{X}}$ or $\bar{T}P_1$, where \bar{T} is the transformation of \mathcal{P} induced by T ; see Chapter 6 of Lehmann and Romano (2005).

where $\Gamma(P_1, P_2)$ denotes the collection of all distributions (over $\mathbb{R}^d \times \mathbb{R}^d$, say; but Kantorovich is working with abstract metric spaces) with given marginals P_1 and P_2 . Kantorovich’s problem clearly constitutes a relaxation of Monge’s problem. The huge advantage of this new formulation is that the class of feasible solutions (the collection $\Gamma(P_1, P_2)$ of distributions with marginals P_1 and P_2) now is convex, so that (3) reduces to a linear optimization problem over a convex set for which Kantorovich develops a powerful duality approach. For the quadratic loss function $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$ and under continuity assumptions, the solutions of Kantorovich’s problem (3) are of the form

$$\gamma = (\text{Identity} \times T) \# P_1$$

with $(\text{Identity} \times T)\mathbf{X} := (\mathbf{X}, T\mathbf{X})$ where the transport T solves Monge’s problem (2).

The topic attracted a renewed surge of interest in the 1990s. Still for the quadratic loss function, Cuesta-Albertos and Matrán (1989) established (under continuity assumptions and the existence of finite second-order moments) the existence of solutions for Monge’s problem. Rüschendorf and Rachev (1990) characterized these solutions in terms of gradients of convex (potential) functions. Brenier (1991), with his celebrated *polar factorization theorem*, independently obtained the same results and, moreover, proved the (a.s.) uniqueness of the solution.

Whether described as in (1), or relaxed into the more general coupling form (3), the so-called Monge-Kantorovich problem remains an optimization problem, though, which only makes sense under densities for which expected transportation costs are finite—under finite variances, thus, for quadratic loss. In a context of nonparametric or semiparametric inference, one clearly would like to avoid such an assumption. This is made possible thanks to a remarkable result by McCann (1995), hereafter *McCann’s theorem*, the nature of which is geometric rather than analytical. Contrary to Monge, Kantorovich, and Brenier, McCann does not require any moment restrictions; his theorem implies that, for any given absolutely continuous distributions P_1 and P_2 over \mathbb{R}^d , there exist convex functions $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ the a.e. gradients³ $\nabla\psi$ of which are pushing P_1 forward to P_2 . Although ψ may not be unique, $\nabla\psi$ is P_1 -a.s. uniquely determined.⁴ Moreover, under the existence of finite moments of order two, $\nabla\psi$ is an L^2 -optimal (in the Monge-Kantorovich sense) transport pushing P_1 forward to P_2 .

We refer to Villani (2003, 2009) for background reading and an authoritative survey of the subject, along with the two volumes by Rachev and Rüschendorf (1998), where the scope is closer to probabilistic and statistical concerns.

3 DISTRIBUTION AND QUANTILE FUNCTIONS, RANKS, AND SIGNS

Distribution functions and their inverses, the quantile functions, along with their empirical counterparts involving ranks and signs, play a central role in probability theory and statistical

³Recall that a convex function is a.e. differentiable.

⁴Meaning that if two convex functions ψ_1 and ψ_2 are such that $\nabla\psi_1 \# P_1 = P_2 = \nabla\psi_2 \# P_1$, then $P_1[\{\mathbf{x} : \nabla\psi_1(\mathbf{x}) \neq \nabla\psi_2(\mathbf{x})\}] = 0$.

decision theory. Many attempts have been made, therefore, to extend the well-understood univariate concepts to higher dimension.

3.1 Ranks in dimension $d = 1$

Before reviewing these attempts, however, an obvious preliminary question arises: what are the properties expected from these concepts? In other terms, what is it that makes them natural and successful tools for statistical inference in a univariate context? Focusing on ranks (hereafter ranks and signs) and their role in testing problems, an immediate answer is *distribution-freeness*: ranks enter the picture in semiparametric statistical models under which the distribution $P_{\boldsymbol{\theta},f}^{(n)}$ of some real-valued observation $\mathbf{X} = (X_1, \dots, X_n)'$, besides a finite-dimensional parameter of interest $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, also depends on the unspecified density $f \in \mathcal{F}_1$ (\mathcal{F}_1 the family of Lebesgue densities over \mathbb{R}) of some unobserved univariate noise $Z_i(\boldsymbol{\theta})$, say. More precisely, models under which $\mathbf{X} \sim P_{\boldsymbol{\theta},f}^{(n)}$ iff some $\boldsymbol{\theta}$ -residuals $Z_1(\boldsymbol{\theta}), \dots, Z_n(\boldsymbol{\theta}) =: \mathbf{Z}^{(n)}(\boldsymbol{\theta})$ are i.i.d. with density f .⁵ In such models, testing the null hypothesis $H_0^{(n)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ (that is, $P_{\boldsymbol{\theta},f}^{(n)} \in \mathcal{P}_{\boldsymbol{\theta}_0}^{(n)} := \{P_{\boldsymbol{\theta}_0,f}^{(n)} | f \in \mathcal{F}_1\}$) reduces to the problem of testing that $Z_1(\boldsymbol{\theta}_0), \dots, Z_n(\boldsymbol{\theta}_0)$ are i.i.d. with unspecified density $f \in \mathcal{F}_1$. Rank tests of $H_0^{(n)}$ are based on the ranks $\mathbf{R}^{(n)}(\boldsymbol{\theta}_0) := (R_1^{(n)}(\boldsymbol{\theta}_0), \dots, R_n^{(n)}(\boldsymbol{\theta}_0))$ of $Z_1(\boldsymbol{\theta}_0), \dots, Z_n(\boldsymbol{\theta}_0)$. These ranks, under $H_0^{(n)}$, are uniform over the $n!$ permutations of $\{1, \dots, n\}$ irrespective of $f \in \mathcal{F}_1$ and rank tests, therefore, are distribution-free.

Ranks, thus, are distribution-free—in other terms, the σ -field they generate is *ancillary*. Ranks are not the only distribution-free statistics, though: for instance, denoting by $Z_{(1/2)}$ the median of $Z_1(\boldsymbol{\theta}_0), \dots, Z_n(\boldsymbol{\theta}_0)$, the signs $\mathbf{S}^{(n)}(\boldsymbol{\theta}_0) := (S_1^{(n)}(\boldsymbol{\theta}_0), \dots, S_n^{(n)}(\boldsymbol{\theta}_0))$, where $S_i^{(n)}(\boldsymbol{\theta}_0)$ is the sign of $Z_i(\boldsymbol{\theta}_0) - Z_{(1/2)}$, are (assuming, for simplicity, that n is even) uniform over the $n(n-1) \dots (n/2-1)$ permutations with repetition of a set of $n/2$ “+1”s and $n/2$ “-1”s. The σ -field generated by the signs thus is ancillary, but a strict sub- σ -field of the σ -field generated by the ranks. Actually, it can be shown (Hallin et al., 2021a, Appendix D.1) that the σ -field of the ranks is *essentially maximal ancillary*—that is, maximal, in the sense of set inclusion and up to sets with Lebesgue measure zero, in the collection of all ancillary σ -fields: intuitively, the ranks are “maximal distribution-free.”

Ancillarity, furthermore, is strongly related, via Basu’s theorems (see, again, Appendix D.1 in Hallin et al. (2021a)), to independence with respect to the order statistic of $Z_1(\boldsymbol{\theta}_0), \dots, Z_n(\boldsymbol{\theta}_0)$, which is minimal sufficient for the family $\{P_{\boldsymbol{\theta},f}^{(n)} | f \in \mathcal{F}_1\}$ (\mathcal{F}_1) and thus contains the relevant information, the whole relevant information, and nothing but the relevant information, about the nuisance f . Accordingly, an essentially maximal ancillary σ -field carries the residual information about the parameter of interest $\boldsymbol{\theta}$. This, which explains the success of univariate rank methods in inference about $\boldsymbol{\theta}$ in the presence of an unspecified f , has been formalized in Hallin and Werker (2003).

⁵Typical 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 $Z_i(\boldsymbol{\theta}) = X_i - \boldsymbol{\theta}'(X_{i-1}, \dots, X_{i-q})$ (where i denotes time and $\boldsymbol{\theta} \in \mathbb{R}^q$; see, e.g., Hallin and Werker (1998)), etc.

Besides their role in statistical decision theory, univariate ranks also are intimately connected to the fundamental probabilistic concepts of distribution and quantile functions. When computed at Z_i , indeed, the empirical distribution function of Z_1, \dots, Z_n takes the form⁶ $F^{(n)}(Z_i) = R_i^{(n)}/(n+1)$; this entirely characterizes $F^{(n)}$, as all other values are obtained from an arbitrary non-decreasing⁷ interpolation of the n -tuple $(Z_1, F^{(n)}(Z_1)), \dots, (Z_n, F^{(n)}(Z_n))$. The emblematic property of $F^{(n)}$ is the Glivenko–Cantelli Theorem connecting it with the underlying population distribution function F , namely,

$$\lim_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} \left(F^{(n)}(z) - F(z) \right) = 0 \text{ or, equivalently, } \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \left(F^{(n)}(Z_i) - F(Z_i) \right) = 0 \text{ a.s.}$$

where the Z_i 's are i.i.d. with distribution function F .

Distribution functions (hence quantile functions) and ranks, thus, cannot be separated: a sound multivariate extension of the traditional concept of ranks should come with an extension of the corresponding concept of distribution (hence quantile) function. The two concepts should be related via some Glivenko–Cantelli property, while the proposed ranks (together, as we shall see, with the signs) should enjoy the maximal ancillarity property of their univariate counterparts.

3.2 Ranks in dimension $d > 1$: a survey

Let us now, in the light of the previous section, briefly review the various concepts of ranks for multivariate data that have been considered in the literature. To facilitate the exposition, let us focus on ranks and their role in testing problems: most existing rank tests belong to one of the following four types.

(i) Componentwise ranks. In line with the traditional copula-transform-related definition of distribution functions in \mathbb{R}^d , componentwise ranks simply rank univariate marginals: see Hodges (1955) for an early example, the monograph by Puri and Sen (1971) for a systematic coverage of classical multivariate analysis problems with independent observations, Hallin et al. (1989) for time series applications. Rather than addressing the tricky problem of ordering \mathbb{R}^d , componentwise ranks 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 coinciding with the chosen coordinates, componentwise ranks are not even asymptotically distribution-free—let alone maximal ancillary; marginal Glivenko–Cantelli properties hold, but the d -tuple of marginal distribution functions fails to characterize the underlying distribution.

(ii) Spatial ranks and signs and related concepts. This class of multivariate ranks and signs includes various very ingenious and elegant concepts, developed by several authors; we refer to Oja (1999), Marden (1999), Oja and Randles (2004), and the monograph by Oja (2010) for details, systematic reviews and exhaustive lists of references. A related concept of quantile has been proposed by Chaudhuri (1996); on this point, see also Serfling (2002). Appealing as they

⁶We divide by $(n+1)$ rather than by n in order for $F^{(n)}$ to take values in the open unit interval $(0,1)$.

⁷The conventional choice is a càdlàg step function interpolation, but any non-decreasing interpolation is equally fine, containing the same information and satisfying the same Glivenko–Cantelli theorem.

are, however, none of them is enjoying distribution-freeness nor is clearly associated with any definition of a distribution function for which a Glivenko–Cantelli property might hold.

(iii) Depth-based ranks. The various concepts of statistical depth were introduced, mainly, as an attempt to define quantile regions and quantile contours in \mathbb{R}^d . Depth itself is a scalar-valued function and the ranks of the depths of i.i.d. residuals accordingly are well defined and distribution-free. Those ranks have been considered in Liu (1992), Liu and Singh (1993), Zuo and He (2006), Zuo and Serfling (2000), among others; see Serfling (2002) for a general introduction on statistical depth, Hallin et al. (2010a) for the related concept of quantile, Lòpez-Pintado and Romo (2012) for functional extensions, Zuo (2018) for a state-of-the art survey in a regression context. Unfortunately, depth-based ranks fail to generate a maximal ancillary σ -field: they carry no directional information and do not satisfy, e.g., Basu’s condition for (essential) maximal ancillarity (Basu, 1959).

(iv) Mahalanobis ranks and signs. These ranks are obtained (Hallin and Paindaveine, 2002) as the ranks $R_i^{(n)}$ of the Mahalanobis moduli $\left[(\mathbf{Z}_i - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) \right]^{1/2}$ of i.i.d. residuals $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ and, denoting by $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ a consistent empirical location vector and a consistent empirical scatter matrix, respectively, the corresponding signs are the unit vectors $\mathbf{S}_i^{(n)} := \hat{\boldsymbol{\Sigma}}^{-1/2} (\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) / \left\| \hat{\boldsymbol{\Sigma}}^{-1/2} (\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) \right\|$; interdirections (Randles, 1989) or hyperplane-based signs (Oja and Paindaveine, 2005) can be used as well. Under ellipticity assumptions, these ranks and signs can be shown to satisfy Basu’s sufficient condition for (essential) maximal ancillarity (Basu, 1959), while the ranks yield an immediate Glivenko–Cantelli result for the radial distribution function F . The Mahalanobis ranks and signs, thus, satisfy all the properties desired for ranks and signs in a multivariate context—for the restricted family of elliptical distributions, though, whereas we would like this to hold for the whole family of absolutely continuous distributions. They have been used, though, quite successfully, in a number of problems, ranging from multivariate location to homogeneity of scatter, VARMA models, and common principal components (Um and Randles, 1998, Hallin and Paindaveine, 2002, 2004, 2006, 2008b, Hallin et al., 2010b, 2013).

A few other concepts have been proposed as well, such as Belloni and Winkler (2011) or Hamel and Kostner (2018), which are related to *cone ordering* and require some subjective (or problem-specific) preliminary choices; they similarly fail to achieve distribution-freeness, hence (DF⁺).

3.3 Center-outward distribution and quantile functions, ranks, and signs

To facilitate the exposition, we assume throughout that \mathbf{Z}_i , $i = 1, \dots, n$ is an i.i.d. sample with distribution P in the family \mathcal{P}_d of Lebesgue-absolutely continuous distributions over \mathbb{R}^d admitting a *nonvanishing* density f (see (4) for a precise definition); the notation Z_i is used for $d = 1$. Denote by \mathbb{S}_d and \mathcal{S}_{d-1} , respectively, the open unit ball and the unit sphere in \mathbb{R}^d .

3.3.1 Center-outward distribution and quantile functions, ranks and signs, $d = 1$

The linear left-to-right ordering of the real line cannot be expected to extend as such to higher dimension, where left and right make no sense. We therefore start with a *center-outward* version of classical univariate distribution functions and ranks.

The value at $z \in \mathbb{R}$ of the traditional distribution function F is the probability $F(z)$ of the half-line $(-\infty, z]$. Define the *center-outward distribution function* as $\mathbf{F}_\pm := 2F - 1$: the value at $z \in \mathbb{R}$ of \mathbf{F}_\pm then is minus the probability of the interval $[z, F^{-1}(1 - F(z))]$ if z is smaller than the median $z_{(1/2)}$ and the probability of the interval $[F^{-1}(1 - F(z)), z]$ if z is larger than $z_{(1/2)}$. Boldface is used in order to stress the nature of \mathbf{F}_\pm as a vector in the unit ball \mathbb{S}_1 : the modulus $\|\mathbf{F}_\pm\|$ then is a probability and $\mathbf{F}_\pm/\|\mathbf{F}_\pm\|$ is a sign, with values in the unit sphere $\mathcal{S}_0 = \{-1, +1\}$. It is easy to see that Z has center-outward distribution function \mathbf{F}_\pm iff $\mathbf{F}_\pm(Z)$ is uniform over \mathbb{S}_1 , the open unit ball in \mathbb{R} ; moreover, $\|\mathbf{F}_\pm\|$ and $\mathbf{F}_\pm/\|\mathbf{F}_\pm\|$ are mutually independent. The regions $\mathbb{C}_\pm(\tau) := \mathbf{F}_\pm^{-1}(\tau\mathbb{S}_1)$, $\tau \in [0, 1)$, are closed, connected, and nested, with P-probability content τ : thus, $\mathbb{C}_\pm(\tau)$ qualifies as a *center-outward quantile region* of order τ , its boundary $\mathbb{C}_\pm(\tau) := \mathbf{F}_\pm^{-1}(\tau\mathcal{S}_0)$ as the corresponding *center-outward quantile contour*, and the inverse \mathbf{Q}_\pm of \mathbf{F}_\pm as a *center-outward quantile function*.

Obviously, \mathbf{F}_\pm is monotonically increasing, hence the gradient of a convex function, mapping \mathbb{R} to \mathbb{S}_1 . In case P has finite second-order moment, McCann's theorem implies that it is the P-a.s. unique⁸ optimal (for quadratic costs) transport pushing P to the uniform over the unit ball.

Traditional ranks, hence the empirical version $F^{(n)}$ of the distribution function, are obtained as the monotone increasing mapping of the sample values Z_1, \dots, Z_n to a grid of the form

$$\mathfrak{G}^{(n)} := \left\{ \frac{1}{n+1}, \dots, \frac{n}{n+1} \right\},$$

yielding ranks $1, \dots, n$. An empirical version $\mathbf{F}_\pm^{(n)}$ of \mathbf{F}_\pm similarly is obtained as a monotone increasing mapping of the sample values, now to the grid

$$2\mathfrak{G}^{(n)} - 1 = \left\{ \frac{\pm 1}{n+1}, \dots, \frac{\pm \lceil n/2 \rceil}{n+1} \right\},$$

along with the origin in case n is odd; this yields *center-outward ranks* $1, \dots, \lceil n/2 \rceil$ (0 in case n is odd) and *center-outward signs* $\{-1, +1\}$.

These new ranks and signs are essentially equivalent to the traditional ones, of which they are simple bijective functions: see the left-hand panel of Figure 1 for a graph. Accordingly, they define the same (essentially maximal ancillary) σ -field. They obviously satisfy a Glivenko–Cantelli property with respect to the center-outward distribution function.

3.3.2 Monge-Kantorovich depth, vector quantiles, and vector ranks, $d \geq 1$

Chernozhukov et al. (2017), in a very general approach, propose measure transportation-based definitions of a broad class of statistical depths, quantile, and ranks. More precisely, for every

⁸For $P \in \mathcal{P}_1$, the Lebesgue-a.e. unique one.

choice of a compactly supported absolutely continuous reference distribution U , they define a *Monge-Kantorovich vector quantile function* as the a.e. unique gradient $\mathbf{Q}_{\text{MK}} = \nabla\psi$ of a convex function ψ pushing U to P and the corresponding *Monge-Kantorovich vector rank function* as the gradient $\mathbf{F}_{\text{MK}} := \nabla\psi^*$ of ψ 's *Legendre-Fenchel convex conjugate*⁹ ψ^* .

Several natural choices are possible for the reference U , which provides much flexibility to the concept. The spherical uniform U_d over the unit ball yields the center-outward quantiles, ranks, and signs described in Section 3.3.3 which, for $d = 1$, reduce to the center-outward concepts described in Section 3.3.1. The Lebesgue uniform $U_{[0,1]^d}$ over the unit (in the canonical basis) hypercube $[0, 1]^d$ yields vector ranks and quantiles that reduce, for $d = 1$ to the classical distribution and quantile functions, respectively. Uniforms over other unit hypercubes, over centered hypercubes $[-1, 1]^d$, or the spherical multinormal are possible as well.

The empirical versions of Monge-Kantorovich vector ranks and quantiles are obtained as empirical optimal transports between sequences $\{\hat{P}^{(n)}\}$ and $\{\hat{U}^{(n)}\}$ of measures converging weakly, as $n \rightarrow \infty$, to P and U , respectively. This, again, allows for much flexibility in the choice of an estimation method. Obvious choices for $\hat{P}^{(n)}$ are the empirical distribution of an i.i.d. sample from P , or smoothed versions thereof. As for $\hat{U}^{(n)}$, since U , as a rule, is known, one may consider U itself or a uniform over some discrete grid of n points in the support of U . That grid can be random (e.g., an i.i.d. sequence simulated from U) or deterministic (see Chapter 9 of Judd (1998)). The transports, accordingly, are either discrete/discrete, discrete/continuous (“semidiscrete”),¹⁰ continuous/discrete, or continuous/continuous; algorithms for each type exist in the measure transportation literature (see, e.g., Peyré and Cuturi (2019) for references). For each case, Chernozhukov et al. (2017) establish weak¹¹ consistency results under the assumption that \mathbf{Q}_{MK} and \mathbf{F}_{MK} are homeomorphisms and the inverse of each other¹², for compactly supported¹³ P and U .

3.3.3 Center-outward distribution and quantile functions, ranks and signs, $d \geq 1$

Hallin et al. (2021a) are revisiting the concept of vector rank with reference distribution the spherical uniform over the unit ball \mathbb{S}^d —from a totally different perspective, however, putting

⁹The *Legendre-Fenchel convex conjugate* ψ^* of ψ is defined as

$$\mathbf{x} \in \mathbb{R}^d \mapsto \psi^*(\mathbf{x}) := \sup_{\mathbf{u} \in \mathbb{R}^d} [\langle \mathbf{u}, \mathbf{x} \rangle - \psi(\mathbf{u})].$$

¹⁰Computationally, this type of transport, yielding (random) tessellation of the support of U , is particularly attractive. Each tile is associated with a unique observed point; a point estimation of the corresponding vector rank can be obtained as a centroid of the tile. Although the tessellation itself is not distribution-free, Ghosal and Sen (2019) show that defining the centroid via an external uniform randomization over each tile restores distribution-freeness.

¹¹Actually, uniform convergence in *outer probability*, see van der Vaart and Wellner (1996).

¹²The set of sufficient conditions (borrowed Theorem 4.14 in Villani (2003)) provided for this, however, requires both P and U to have bounded densities, a condition which is not satisfied when the reference distribution is the spherical uniform U_d over the unit ball (the density of U_d is unbounded at the origin).

¹³Assuming P to be compactly supported, hence having finite second-order moments, is unfortunate when defining concepts of ranks and quantiles.

emphasis on the role of the vector rank function as a distribution function extending the univariate concept of center-outward distribution function introduced in Section 3.3.1 and satisfying, along with the empirical transport from the sample distribution function to a regular grid of \mathbb{S}^d , the properties developed in Section 3.1. This justifies the terminology *center-outward distribution and quantile functions* (with the notation \mathbf{F}_\pm and \mathbf{Q}_\pm) instead of *Monge-Kantorovich vector ranks and vector quantiles*.

For simplicity, we limit our description to distributions P in the family \mathcal{P}_d of all Lebesgue-absolutely continuous distributions over \mathbb{R}^d with probability density f satisfying the “nonvanishing density” condition

$$\text{for all } D \in \mathbb{R}^+, \text{ there exist constants } \Lambda_{D;f} \text{ and } \lambda_{D;f} \in (0, \infty) \quad (4)$$

$$\text{such that } \lambda_{D;f} \leq f(\mathbf{x}) \leq \Lambda_{D;f} \text{ for all } \mathbf{x} \in (D\mathbb{S}_d).$$

The support of $P \in \mathcal{P}_d$ thus is \mathbb{R}^d ; Hallin et al. (2021a), however, also consider absolutely continuous distributions supported on strict subsets of \mathbb{R}^d .

More precisely, assuming that $\mathbf{Z} \sim P \in \mathcal{P}_d$, the *center-outward distribution function* \mathbf{F}_\pm of P (of \mathbf{Z}) can be defined as the a.e. unique gradient of a convex function pushing P forward to U_d . A regularity result by Figalli (2018) implies that \mathbf{F}_\pm actually is unique and a homeomorphism between the punctured ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$ and $\mathbb{R}^d \setminus \mathbf{F}_\pm^{-1}(\{\mathbf{0}\})$; over these domains, thus, \mathbf{F}_\pm has a well-defined and homeomorphic inverse \mathbf{Q}_\pm which is easily extended as a multivalued function with domain the entire ball by letting $\mathbf{Q}_\pm(\mathbf{0}) := \mathbf{F}_\pm^{-1}(\{\mathbf{0}\})$; \mathbf{Q}_\pm naturally qualifies as a *center-outward quantile function*, $\mathbf{Q}_\pm(\mathbf{0})$ as a *center-outward median*.¹⁴ Note that Figalli’s result establishes, under a slightly weaker form (the continuity of \mathbf{Q}_\pm may not hold at the origin) and for $P \in \mathcal{P}_d$, the regularity properties that Chernozhukov et al. (2017) need to assume for their consistency results.

The empirical counterparts $\mathbf{F}_\pm^{(n)}$ and $\mathbf{Q}_\pm^{(n)}$ of \mathbf{F}_\pm and \mathbf{Q}_\pm are obtained via an optimal discrete transport of the sample distribution to a deterministic regular grid $\mathfrak{G}^{(n)}$ of \mathbb{S}_d —an optimal pairing between the n sample points $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ (possibly, residuals) and the gridpoints. That grid is obtained as follows. Factorizing n into $n_R n_S + n_0$ where $n_0 < \min(n_R, n_S)$, consider a “regular array” $\mathfrak{S}_{n_S} := \{\mathbf{s}_1^{n_S}, \dots, \mathbf{s}_{n_S}^{n_S}\}$ of n_S points on \mathcal{S}_{d-1} (see the comment below); $\mathfrak{G}^{(n)}$ consists of the $n_R n_S$ points of the form $(r/(n_R + 1))\mathbf{s}_s^{n_S}$ for $r = 1, \dots, n_R$ and $s = 1, \dots, n_S$, along with n_0 copies of the origin in case $n_0 \neq 0$; see the right-hand panel of Figure 1 for a graph in dimension two. By “regular” we mean “as uniform as possible”, in the sense, for example, of 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)). The only mathematical requirements (in asymptotic statements) is the convergence to infinity of both n_R and n_S as $n \rightarrow \infty$ and the weak convergence, as $n_S \rightarrow \infty$, of the uniform discrete distribution over \mathfrak{S}_{n_S} to the uniform distribution over \mathcal{S}_{d-1} .

Call *center-outward rank* of \mathbf{Z}_i the integer $R_{i;\pm}^{(n)} := (n_R + 1) \|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i)\|$ (with value in $\{1, \dots, n_R\}$ or $\{0, \dots, n_R\}$ according as $n_0 = 0$ or not) and *center-outward sign* of \mathbf{Z}_i the unit

¹⁴That median region, moreover, is a convex compact subset of \mathbb{R}^d , with Hausdorff dimension at most $d - 1$; for $1 \leq d \leq 3$, it consists of a single point.

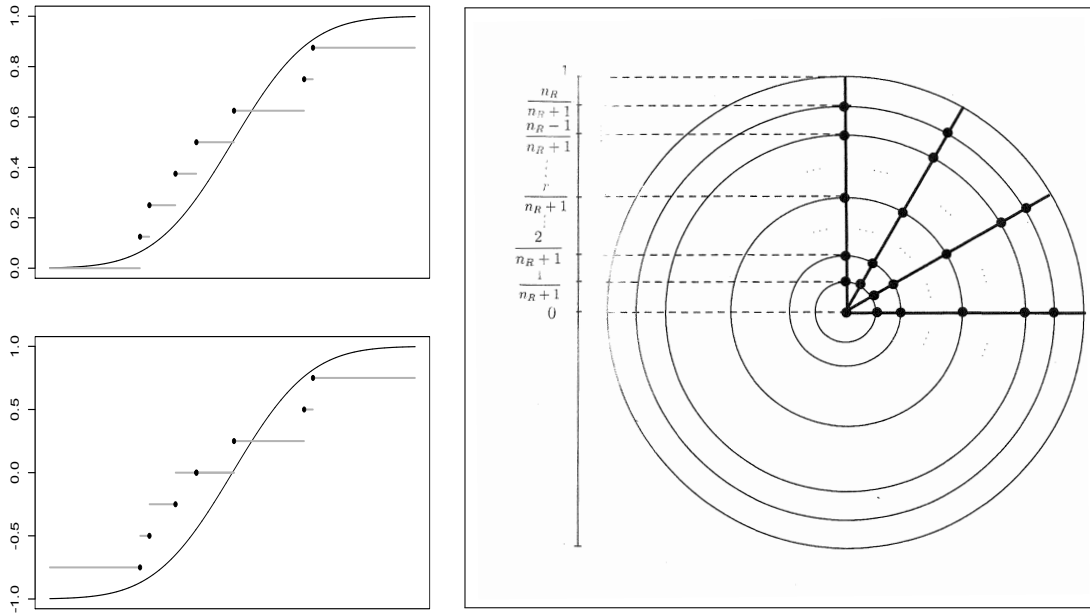


Figure 1: A classical distribution function F and its empirical counterpart $F^{(n)}$ for $n = 7$ (top left panel), along with (bottom left panel) their center-outward versions \mathbf{F}_\pm and $\mathbf{F}_\pm^{(n)}$, the latter with left-continuous piecewise constant interpolation on the left-hand side of the (empirical) median, right-continuous piecewise constant interpolation on the right-hand side of the median; a regular grid of $n = n_R n_S$ points over \mathbb{S}_2 (right panel).

vector $\mathbf{S}_{i;\pm}^{(n)} := \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i) / \|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i)\|$ for $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i) \neq \mathbf{0}$; for $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i) = \mathbf{0}$ (\mathbf{Z}_i then is an empirical median), put $\mathbf{S}_{i;\pm}^{(n)} = \mathbf{0}$.

Some desirable finite-sample properties, such as strict independence between the ranks and the signs, only hold for $n_0 = 0$ or 1 , due to the fact that the mapping from the sample to the grid is no longer injective for $n_0 \geq 2$. This, which has no asymptotic consequences (since the number n_0 of tied values involved is $o(n)$ as $n \rightarrow \infty$), is easily taken care of by an appropriate tie-breaking modification of $\mathfrak{G}^{(n)}$ (see Section 2.2 of Hallin et al. (2021a)). The modified grid no longer has multiple points, and the optimal pairing between the sample and the grid is bijective; the n_0 smallest ranks, however, take the non-integer “mid-rank” value $1/2$. Below, we tacitly assume that either $n_0 = 0$ or $n_0 > 0$ and the tie-breaking device is adopted.

Hallin et al. (2021a) then establish the following main properties:

- (i) \mathbf{F}_{\pm} is a probability integral transformation of \mathbb{R}^d : namely, $\mathbf{Z} \sim P$ iff $\mathbf{F}_{\pm}(\mathbf{Z}) \sim U_d$; by construction, $\|\mathbf{F}_{\pm}(\mathbf{Z})\|$ is uniform over the interval $[0, 1]$, $\mathbf{F}_{\pm}(\mathbf{Z}) / \|\mathbf{F}_{\pm}(\mathbf{Z})\|$ uniform over the sphere \mathcal{S}_{d-1} , and they are mutually independent;
- (ii) for $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ i.i.d. with distribution $P \in \mathcal{P}_d$ and center-outward distribution function \mathbf{F}_{\pm} , $(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \dots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)}))$ is uniformly distributed over the $n!$ permutations of the grid-points in \mathfrak{G}_n ;
- (iii) the n -tuple $(R_{1;\pm}^{(n)}, \dots, R_{n;\pm}^{(n)})$ of center-outward ranks and the n -tuple $(\mathbf{S}_{1;\pm}^{(n)}, \dots, \mathbf{S}_{n;\pm}^{(n)})$ of center-outward signs are mutually independent;
- (iv) the n -tuple $(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_1^{(n)}), \dots, \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_n^{(n)}))$ is *strongly essentially maximal ancillary*;
- (v) the center-outward quantile regions $\mathbf{Q}_{\pm}(\tau\mathcal{S}_d)$, $\tau \in [0, 1)$ are closed, connected and nested, with probability content τ ;
- (vi) (Glivenko–Cantelli) $\max_{1 \leq i \leq n} \|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i) - \mathbf{F}_{\pm}(\mathbf{Z}_i)\| \rightarrow 0$ *a.s.* as $n \rightarrow \infty$.

These properties nicely extend to dimension 2 and higher the properties (see Section 3.3.1) of the univariate concepts. In particular, maximal ancillarity guarantees that center-outward ranks and signs carry all the available “distribution-free information.” Some of these properties also may hold for other particular cases of Monge-Kantorovich vector ranks and empirical ranks: for instance, vector ranks based on transports to the unit cube also constitute probability integral transformations and their empirical versions, when associated with a deterministic grid over the unit cube, also enjoy maximal ancillarity. But the corresponding quantile regions are not well defined, or make little sense (see Section 3.4); the resulting vector ranks very much depend on the coordinate system (their behavior under orthogonal transformations, for instance, is quite complicated, whereas center-outward distribution functions are nicely equivariant). While keeping the Glivenko–Cantelli behavior of their center-outward counterparts, empirical center-outward ranks and signs based on a grid of randomly generated spherical uniform points no longer are mutually independent. Empirical vector ranks resulting from semi-discrete transportation

are not distribution-free (unless an additional randomization is performed: see Ghosal and Sen (2019)), etc.

Empirical center-outward distribution functions, however, only are defined at the observations; empirical center-outward quantile contours are just a collection of discrete points. Hallin et al. (2021a) therefore propose a smooth interpolation $\bar{\mathbf{F}}_{\pm}^{(n)}$ of $\mathbf{F}_{\pm}^{(n)}$ within the class of center-outward distribution functions¹⁵—the d -dimensional version of a smooth monotone nondecreasing interpolation of the n points $(Z_i, \mathbf{F}_{\pm}^{(n)}(Z_i))$ for $d = 1$. The proposed interpolation $\bar{\mathbf{F}}_{\pm}^{(n)}$, based on *Moreau envelopes* and *Yosida regularization* (Yosida, 1964), is the gradient of some convex function, carries the same information as $\mathbf{F}_{\pm}^{(n)}$, and satisfies a Glivenko–Cantelli property under sup form:

$$\sup_{\mathbf{z} \in \mathbb{R}^d} \left\| \bar{\mathbf{F}}_{\pm}^{(n)}(\mathbf{z}) - \mathbf{F}_{\pm}(\mathbf{z}) \right\| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

Figure 2 provides the picture of the smooth interpolation $\bar{\mathbf{F}}_{\pm}^{(n)}$ of the empirical center-outward distribution function obtained from $n = 20000$ i.i.d. observations from a mixture of three bivariate Gaussian distributions (we refer to Hallin et al. (2021a) for a detailed description of the mixture). Notice the quantile contours (in red) and the sign curves (in blue).

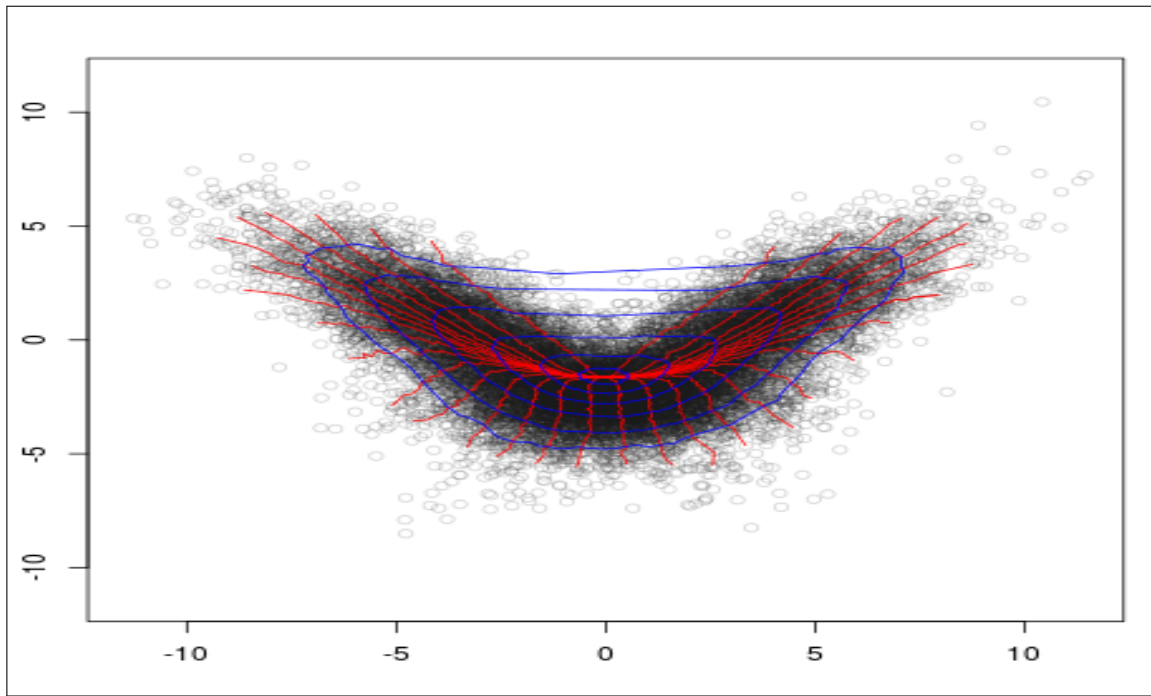


Figure 2: Empirical center-outward quantile contours and sign curves ($n = 20000$) for a mixture of three bivariate Gaussian distributions.

¹⁵That class is the collection of all gradients of convex functions mapping \mathbb{R}^d to \mathbb{S}_d , characterized by the property of *cyclical monotonicity*.

3.4 Unit cube or unit ball?

The Lebesgue uniform over the unit cube $[0, 1]^d$ and the spherical uniform over the unit ball \mathcal{S}_d yield transformations \mathbf{F}_{MK} and \mathbf{F}_{\pm} reducing, for $d = 1$, to the traditional distribution function F and the center-outward distribution function $2F - 1$, respectively. Thus, both can be considered as extending the univariate concept of a distribution function.

An inconvenient feature of unit cubes, however, is that there are many of them: besides $[0, 1]^d$, constructed over the canonical coordinate system, all orthogonal transformations of $[0, 1]^d$ are equally “natural.” The corresponding vector ranks \mathbf{F}_{KM} and vector quantiles \mathbf{Q}_{KM} all carry the same information but their equivariance properties at best are unclear. While each orthogonal transformation provides a distinct unit cube and distinct vector rank/quantile maps with poor equivariance properties, the unit ball and U_d are orthogonally invariant, and \mathbf{F}_{\pm} and \mathbf{Q}_{\pm} are nicely equivariant.

Poor equivariance properties, however, are not the main drawback of the vector ranks \mathbf{F}_{KM} and vector quantiles \mathbf{Q}_{KM} associated with $U_{[0,1]^d}$: the very interpretation of \mathbf{Q}_{KM} as a quantile is problematic. Indeed, $\mathbf{Q}_{\text{KM}}(\mathbf{u})$ is characterized by a d -dimensional *quantile order* $\mathbf{u} = (u_1, \dots, u_d)$, where each u_j has a probabilistic interpretation that strongly depends on the coordinate system (the choice of the reference unit cube) and does not yield straightforward notions of quantile regions and contours. To palliate this, Chernozhukov et al. (2017) define the Monge-Kantorovich depth region of order τ as the image by \mathbf{Q}_{KM} of the depth region with probability content τ of $U_{[0,1]^d}$ —viz. the region of all points in the unit cube with halfspace depth at least $D = D(\tau)$, where $D(\tau)$ is such that the $U_{[0,1]^d}$ -probability of that region is τ . Monge-Kantorovich depth regions and contours of order τ are to be interpreted as quantile regions and contours of order τ .¹⁶ As a consequence, the concepts of vector quantile map and those of quantile region and contour, if not totally unrelated, are somewhat disconnected.

When the unit ball and U_d are considered instead of the unit cube and $U_{[0,1]^d}$, the concept of (center-outward) quantile contour of order $\tau \in [0, 1)$ and the concept of depth contour of order τ (now the images by \mathbf{Q}_{\pm} of U_d 's depth contour of order τ —viz. the hypersphere centered at the origin with radius τ , hence probability content τ)—coincide. This not only fully reconciles depth and quantiles, it eventually achieves the ultimate goal of the theory of statistical depth.

Bibliographic note

Galichon and Henry (2012) and Ekeland et al. (2012) with their concept of *multivariate comonotonicity* were early forerunners in the developments that eventually led to the concepts of center-outward distribution functions, ranks, signs, and quantiles. Their objective was a theory of coherent risk measurement, though, not a contribution to statistical decision theory. Based on similar ideas, the definitions of Section 3.3.2 were developed in Carlier et al. (2016) with the objective of extending to the multiple-output context the classical theory of quantile regression and in Chernozhukov et al. (2017) where the focus is on depth rather than distribution functions and

¹⁶The shape of that region, again, strongly depends on the choice of the reference unit cube; its probability content (its order τ as a quantile region) does not and is a monotone decreasing function of D .

asymptotics (weak consistency rather than Glivenko–Cantelli) requiring compactly supported distributions, hence finite second-order moments. This triggered Faugeras and Rüschendorf (2017), who are bypassing the compact support assumption by performing a copula transform prior to the transport to the reference distribution U . That copula transform and the resulting ranks and quantiles, however, strongly depend on the choice of the coordinate system—a choice that is often irrelevant. The first sections of Hallin et al. (2021a), introducing center-outward distribution and quantile functions, insisting on avoiding compact support and finite moment assumptions, and emphasizing the role of center-outward ranks as fundamental statistical tools, were made available as Hallin (2017). Boeckel et al. (2018) also are revisiting, under the terminology *Brenier distribution function*, the concepts of Chernozhukov et al. (2017) and Hallin (2017). Their empirical Brenier distribution function is obtained as a discrete transport from the sample to a grid of i.i.d. simulated observations from some arbitrary reference distribution U . Under the assumptions of compact supports (finite second-order moments) and continuous Brenier distribution functions, they establish the Glivenko–Cantelli property for any interpolation by a gradient of convex function of their empirical Brenier mapping; they do not provide, however, any constructive way to produce such an interpolation. Such a construction was proposed by del Barrio et al. (2018), which eventually was merged with Hallin (2017) to form Hallin et al. (2021a). The continuity of vector quantile and center-outward quantile functions is an essential issue if meaningful quantile or depth contours and regions are to be defined. In Chernozhukov et al. (2017) and Boeckel et al. (2018), that continuity is an assumption. Solicited for a reference on the regularity of the solutions of Monge–Ampère equations, Alessio Figalli kindly provided (Figalli, 2018) an original proof of the continuity of the center-outward distribution and quantile functions of distributions with non-vanishing densities. His result later on was extended by del Barrio et al. (2020) to distributions with more general supports. In Hallin et al. (2021a), the continuity of \mathbf{F}_\pm and \mathbf{Q}_\pm thus is no longer an assumption but a result.

4 APPLICATIONS IN STATISTICAL INFERENCE

4.1 Center-outward rank tests and R-estimators

The most natural applications of ranks in statistical inference is the construction of rank tests and R-estimators of the parameter of interest $\boldsymbol{\theta}$ in semiparametric models under which the density f of some driving noise remains unspecified in some family \mathcal{F} of densities.

Of daily practice, in this context, is the *pseudo- or quasi-Gaussian* approach. Gaussian tests indeed remain asymptotically valid (i.e., have asymptotic size less than or equal to the nominal size α) for any $f \in \mathcal{F}$ with finite fourth-order moments. One should be aware, however, that this asymptotic validity is far from being uniform with respect to f . Denoting by $\mathbf{P}_{\boldsymbol{\theta};f}^{(n)}$ the distribution under parameter $\boldsymbol{\theta}$ and density f of a sequence $\phi^{(n)}$ of pseudo-Gaussian tests of the hypothesis $H_0^{(n)} := \{\mathbf{P}_{\boldsymbol{\theta}_0;f}^{(n)} \mid f \in \mathcal{F}\}$,

$$\lim_{n \rightarrow \infty} \mathbf{P}_{\boldsymbol{\theta}_0;f}^{(n)} \left[\phi^{(n)} = 1 \right] \leq \alpha \quad \text{for any } f \in \mathcal{F}. \quad (5)$$

This is not sufficient and, for $\phi^{(n)}$ to be asymptotically valid, the stronger condition

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \mathbb{P}_{\theta_0, f}^{(n)} \left[\phi^{(n)} = 1 \right] \leq \alpha, \quad (6)$$

which in general does not follow from (5), is required. When, however, $\phi^{(n)}$ is rank-based, its distribution-freeness under the null automatically implies that (6) is satisfied.

A key role in the construction of rank tests is played by the so-called Hájek representation results establishing, under exchangeability, the asymptotic equivalence (usually, in quadratic mean) between a rank-based statistic and a random variable that no longer involves ranks and satisfies the conditions for standard asymptotic central-limit¹⁷ behavior. Hájek representation results can be used to obtain the asymptotic distribution of a given rank-based test statistic and, via Le Cam's third lemma, the power of the resulting tests against contiguous alternatives (see Chapter VI of Hájek and Šidák (1967)). Starting with Hallin and Puri (1994), Hájek representation results also have been used in the rank-based reconstruction of *central sequences* yielding locally and asymptotically optimal rank tests in locally asymptotically normal (LAN) families.

4.1.1 Multiple-output regression models and MANOVA

Rank tests for single-output linear models (this includes multiple regression, two-sample location, and ANOVA models) have been thoroughly studied in the literature: see Hájek and Šidák (1967) or Puri and Sen (1985). The key role in that literature is played by a Hájek representation result for square-integrable score functions with bounded variation¹⁸ $J : \mathbb{R} \rightarrow \mathbb{R}$ (earlier asymptotic results mainly could handle bounded score functions J). Let $Z_1^{(n)}, \dots, Z_n^{(n)}$ denote a triangular array of i.i.d. variables $Z_i^{(n)}$ with distribution function F and ranks $R_1^{(n)}, \dots, R_n^{(n)}$, and let $c_1^{(n)}, \dots, c_n^{(n)}$, with mean $\bar{c}^{(n)}$, be a triangular array of regression constants satisfying the classical *Noether conditions*). Then, rank statistics of the form

$$\mathcal{T}_{\mathbf{c}}^{(n)} := \left(\sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)})^2 \right)^{-1/2} \sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)}) J \left(\frac{R_i^{(n)}}{n+1} \right) \quad (7)$$

(in Hájek's terminology, *approximate-score linear rank statistics*) admit the asymptotic representation

$$T_{\mathbf{c}}^{(n)} := \left(\sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)})^2 \right)^{-1/2} \sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)}) J(F(Z_i^{(n)})),$$

in the sense that $n^{1/2} \left(\mathcal{T}_{\mathbf{c}}^{(n)} - T_{\mathbf{c}}^{(n)} \right)$ tends to zero in quadratic mean (hence in probability) as $n \rightarrow \infty$: the statistic $\mathcal{T}_{\mathbf{c}}^{(n)}$, thus, is a rank-based, hence distribution-free reconstruction of $T_{\mathbf{c}}^{(n)}$ (which is an *oracle* rather than a statistic since it involves the unspecified F).

¹⁷Rank-based statistics, indeed, typically involve sums of non-mixing sequences of non-independent summands for which traditional central-limit results do not apply.

¹⁸Recall that a function from \mathbb{R} to \mathbb{R} has bounded variation iff it is the difference between two monotone nondecreasing functions.

The same reconstruction, based on center-outward ranks and signs, for multiple-output regression and MANOVA requires a new type of asymptotic representation for the approximate-score center-outward rank statistics (notation of Section 3.3.3)

$$\begin{aligned}\tilde{\mathbf{T}}_{\mathbf{c}}^{(n)} &:= \left(\sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)})^2 \right)^{-1/2} \sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)}) \mathbf{J}(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})) \\ &= \left(\sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)})^2 \right)^{-1/2} \sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)}) \mathbf{J} \left(\frac{R_{i;\pm}^{(n)}}{n_R + 1} \mathbf{S}_{\pm;i}^{(n)} \right)\end{aligned}$$

with score functions $\mathbf{J} : \mathbb{S}_d \rightarrow \mathbb{R}^d$ such that $\int_{\mathbb{S}_d} \mathbf{J}'(\mathbf{u}) \mathbf{J}(\mathbf{u}) dU_d < \infty$ and asymptotic representations of the form $n^{1/2} \left(\tilde{\mathbf{T}}_{\mathbf{c}}^{(n)} - \mathbf{T}_{\mathbf{c}}^{(n)} \right) = o_{\text{q.m.}}(1)$ as $n \rightarrow \infty$ with

$$\mathbf{T}_{\mathbf{c}}^{(n)} := \left(\sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)})^2 \right)^{-1/2} \sum_{i=1}^n (c_i^{(n)} - \bar{c}^{(n)}) \mathbf{J}(\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})); \quad (8)$$

random vectors of the form (8), indeed, appear in central sequences for multiple-output linear models. Hallin et al. (2020a) first establish that multivariate extension of Hájek's result, then use it in the reconstruction of fully distribution-free optimal tests for multiple-output regression and MANOVA.

4.1.2 VAR and VARMA models

LAN for time-series models such as VAR and VARMA yield central sequences of the *serial* type, with scores involving lagged residual values (Garel and Hallin, 1995). In order to construct locally optimal R-estimators and tests based on center-outward ranks and signs, Hallin et al. (2020b,c) therefore also first establish asymptotic representation results, now for serial matrix-valued center-outward rank statistics (to be interpreted as rank-based cross-correlation matrices).

Consider a matrix-valued function $\mathbf{a} : \mathbb{S}_d \times \mathbb{S}_d \rightarrow \mathbb{R}^{d \times d}$ and two continuous *score functions* \mathbf{J}_1 and \mathbf{J}_2 from \mathbb{S}_d to \mathbb{R}^d , define, for $1 \leq i \leq n-1$,

$$\tilde{\mathbf{\Gamma}}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)} := (n-i)^{-1} \sum_{t=i+1}^n \mathbf{a}(\mathbf{F}_{\pm,t}^{(n)}, \mathbf{F}_{\pm,t-i}^{(n)}) \quad \text{and} \quad \mathbf{\Gamma}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)} := (n-i)^{-1} \sum_{t=i+1}^n \mathbf{J}_1(\mathbf{F}_{\pm,t}) \mathbf{J}_2'(\mathbf{F}_{\pm,t-i}),$$

where $\mathbf{F}_{\pm,t}^{(n)}$ is the value at $\mathbf{Z}_t^{(n)}$ of the empirical center-outward distribution function computed from the i.i.d. n -tuple $\mathbf{Z}_1^{(n)}, \dots, \mathbf{Z}_n^{(n)}$ with center-outward distribution function \mathbf{F}_{\pm} and $\mathbf{F}_{\pm,t}$ stands for $\mathbf{F}_{\pm}(\mathbf{Z}_t^{(n)})$. Assume that \mathbf{a} satisfies

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\| \text{vec} \left(\mathbf{a}(\mathbf{F}_{\pm,2}^{(n)}, \mathbf{F}_{\pm,1}^{(n)}) - \mathbf{J}_1(\mathbf{F}_{\pm,2}) \mathbf{J}_2'(\mathbf{F}_{\pm,1}) \right) \right\|^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (9)$$

let $\mathbf{m}^{(n)} := \mathbb{E}[\mathbf{a}(\mathbf{F}_{\pm,2}^{(n)}, \mathbf{F}_{\pm,1}^{(n)})]$ and $\mathbf{m} := \mathbb{E}[\mathbf{J}_1(\mathbf{F}_{\pm,2}) \mathbf{J}_2'(\mathbf{F}_{\pm,1})]$. Clearly, $\mathbf{m}^{(n)}$, which does not depend on the density f of \mathbf{Z}_1 , is centering $\tilde{\mathbf{\Gamma}}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)}$, while $\mathbf{m} = \mathbf{m}_f$, which depends on f

through \mathbf{F}_\pm , is centering $\mathbf{\Gamma}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)}$; none of them depends on i . Then, for any $i \in \mathbb{N}$,

$$(n-i)^{1/2} \text{vec} \left(\mathbf{\Gamma}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)}(\boldsymbol{\theta}_0) - \mathbf{m}^{(n)} - \mathbf{\Gamma}_{i, \mathbf{J}_1, \mathbf{J}_2}^{(n)}(\boldsymbol{\theta}_0) + \mathbf{m} \right) = o_{\text{q.m.}}(1) \quad (10)$$

as $n \rightarrow \infty$. This asymptotic representation is used, in Hallin et al. (2020c), to construct locally asymptotically optimal center-outward rank tests for VAR parameters and the identification of VAR order. Hallin et al. (2020b) consider a weaker form of (10), with *spherical* score functions of the form $\mathbf{J}_\ell(\mathbf{F}_\pm) = J_\ell(\|\mathbf{F}_\pm\|)\mathbf{F}_\pm/\|\mathbf{F}_\pm\|$, in the definition of locally asymptotically optimal center-outward R-estimators in VARMA models.

4.1.3 Vector independence

The problem of consistent (against any form of dependence) testing of the hypothesis of independence between random vectors has a long history, and many attempts have been made to construct distribution-free solutions; see the introduction sections of Shi et al. (2021a) and Deb et al. (2020) for references. The measure-transportation-based concepts introduced in Chernozhukov et al. (2017) and Hallin (2017) are bringing such a solution and have triggered no less than five papers of increasing generality—Ghosal and Sen (2019), Deb and Sen (2019), Shi et al. (2021a), Shi et al. (2020), and Deb et al. (2020)—proposing, in a very short span of time, distribution-free solutions to that important problem.

Ghosal and Sen (2019), Deb and Sen (2019), and Shi et al. (2021a) (posted almost simultaneously on arXiv in 2019) came first. Starting from the observation that the consistent tests available in the literature—among them, the kernel-based tests (Gretton et al. (2005b), Gretton et al. (2005c), Gretton et al. (2005a), Gretton et al. (2008), Sejdinovic et al. (2013)) and the distance covariance-based tests (Bakirov et al., 2006, Székely et al., 2007, Székely and Rizzo, 2009)—are not distribution-free (not even asymptotically so), which is methodologically regrettable and computationally costly, the three papers propose, for the same problem, consistent tests based on the concepts of ranks introduced by Chernozhukov et al. (2017) and/or Hallin (2017).

Before describing their tests, Ghosal and Sen (2019) establish, for vector ranks and quantiles, several general results paralleling, completing, or extending those by Chernozhukov et al. (2017) and Hallin et al. (2021a) (the latter, however, are focusing on the reference distribution $\mathbf{U} = \mathbf{U}_d$ while Ghosal and Sen (2019) consider general \mathbf{U}). In particular, they show that if \mathbf{P} and \mathbf{U} (notation of Section 3.3.2) are absolutely continuous with bounded convex supports, with the density f of \mathbf{P} satisfying (4) over its support and \mathbf{U} having bounded density, then the vector rank and quantile maps \mathbf{F}_{MK} and \mathbf{Q}_{MK} are homeomorphisms. The pole of \mathbf{U}_d at the origin thus explains the possibility of a singularity of \mathbf{Q}_\pm at $\mathbf{0}$ in the corresponding results of Figalli (2018) and del Barrio et al. (2020). Ghosal and Sen (2019) also establish a Glivenko–Cantelli property for homeomorphic vector rank maps \mathbf{F}_{MK} . That result reinforces the weak consistency result in Chernozhukov et al. (2017) which, moreover, requires finite moments of order two. But it does not imply the Glivenko–Cantelli result in Hallin et al. (2021a): indeed, \mathbf{F}_\pm , because of the

possible singularity of \mathbf{Q}_\pm at $\mathbf{0}$, needs not be a homeomorphism on \mathbb{S}_d hence satisfy Ghosal and Sen’s necessary (their Remark 4.2) condition.¹⁹ Finally, a few steps are made in the direction of deriving the local uniform rates of convergence of the empirical vector quantile and rank functions—a difficult problem that deserves further attention.

The main applications of the Ghosal and Sen (2019) paper is the construction of rank-based tests for testing independence between random vectors and two-sample goodness of fit (see Section 4.1.4). The test statistics they are proposing are of the Cramér-von Mises type, with vector rank maps (based on the uniform over $[0, 1]^d$ reference distribution $U = U_{[0,1]^d}$) substituting the traditional empirical distribution functions. More precisely, denote by $\mathbf{F}_{\text{MK};\mathbf{X}}^{(n)}$, $\mathbf{F}_{\text{MK};\mathbf{Y}}^{(n)}$, and $\mathbf{F}_{\text{MK}}^{(n)}$ the empirical vector rank maps computed from $(\mathbf{X}_1, \dots, \mathbf{X}_n)$, $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ and $(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, respectively, where $\mathbf{Z}_i = (\mathbf{X}'_i, \mathbf{Y}'_i)'$, $i = 1, \dots, n$ are i.i.d. copies of $\mathbf{Z} = (\mathbf{X}', \mathbf{Y}')'$, with $d^{\mathbf{X}}$ -dimensional \mathbf{X} , $d^{\mathbf{Y}}$ -dimensional \mathbf{Y} , hence d -dimensional \mathbf{Z} ($d = d^{\mathbf{X}} + d^{\mathbf{Y}}$): the test statistic is²⁰

$$T_n := \frac{1}{n} \sum_{i=1}^n \left\| \mathbf{F}_{\text{MK}}^{(n)} \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} - \begin{pmatrix} \mathbf{F}_{\text{MK};\mathbf{X}}^{(n)}(\mathbf{X}_i) \\ \mathbf{F}_{\text{MK};\mathbf{Y}}^{(n)}(\mathbf{Y}_i) \end{pmatrix} \right\|^2.$$

This test statistic is distribution-free under the null hypothesis of independence between \mathbf{X} and \mathbf{Y} . Finite-sample critical values in principle can be obtained via permutational techniques which, however, are computationally heavy. The test rejecting independence for large values of T_n is shown to be consistent against all forms of dependence. No asymptotic distribution is provided.

Another heuristic way of constructing distribution-free rank tests consists in replacing, in some test statistic chosen in the literature, the observations with their ranks. Starting from the distance covariance-based tests popularized by Gábor Székely and his coauthors, this is the strategy adopted in Deb and Sen (2019) (using general vector ranks $\mathbf{F}_{\text{MK}}^{(n)}$) and Shi et al. (2021a) (using the center-outward ranks and signs $\mathbf{F}_\pm^{(n)}$). Explicit expressions are somewhat cumbersome, and we do not reproduce them here. The resulting tests are fully distribution-free and consistent, and both papers are deriving the asymptotic null distributions of the test statistics, so that computationally costly permutational methods²¹ can be avoided in practical implementation. The methods of proof in the two papers are, however, quite different.²² Shi et al. (2021a) actually establish for their center-outward distance covariance statistic a Hájek representation result with a so-called “oracle distance covariance statistic,” the limiting distribution of which they obtain by combining U-statistic techniques and a combinatorial non-central limit theorem. The same limiting distribution—that of a linear combination of independent chi-square variables with one degree of freedom, the coefficients of which are the eigenvalues of some integral equation—

¹⁹See, however, their Section 4.2 for a discussion.

²⁰If empirical vector ranks are obtained via grids, note that the grid for \mathbf{Z} (n points) cannot be the product of the grids for \mathbf{X} and \mathbf{Y} (n^2 points).

²¹If permutational critical values were to be used, the benefits of considering ranks are unclear, as exact permutational critical values can be obtained for the original distance covariance test statistic as well.

²²Moreover, different reference distributions may require different tools: for instance, the hypercube $[0, 1]^{d^{\mathbf{X}}+d^{\mathbf{Y}}}$ conveniently factorizes into $[0, 1]^{d^{\mathbf{X}}} \times [0, 1]^{d^{\mathbf{Y}}}$, while $\mathbb{S}_{d^{\mathbf{X}}+d^{\mathbf{Y}}}$ and $\mathbb{S}_{d^{\mathbf{X}}} \times \mathbb{S}_{d^{\mathbf{Y}}}$ are distinct.

is derived by Deb and Sen (2019) using characteristic function methods. Simulation studies in both papers yield quite encouraging finite-sample performance, particularly under mixtures distributions and heavy tails.

Shi et al. (2020) are revisiting yet the same problem. However, instead of heuristically replacing the observations with their multivariate ranks in some given classical test statistic, they establish a general Hájek representation result and exploit it in order to obtain center-outward rank-based dependence measures asymptotically reconstructing the broad class of *generalized symmetric covariances* proposed by Weihs et al. (2018). Besides distance covariance concepts, that class also contains, among many others, multivariate versions of Hoeffding’s D based on marginal ordering (see Weihs et al. (2018)) and projection-averaging (see Zhu et al. (2017), of Blum–Kiefer–Rosenblatt’s R , and Bergsma–Dassios–Yanagimoto’s τ^* (see Kim et al. (2020)). As in Hallin et al. (2020a) and Hallin et al. (2020c), their Hájek representation involves *scores* yielding, for instance, Spearman (Wilcoxon), van der Waerden (Gaussian score), or sign test score versions of center-outward rank-based generalized symmetric covariances. By allowing the replacement, in asymptotic properties, of the empirical center-outward $\mathbf{F}_{\pm}^{(n)}$ with its population counterpart \mathbf{F}_{\pm} , this Hájek representation also simplifies the derivation of asymptotic null distributions and, via a nontrivial use of Le Cam’s third lemma for non-normal limiting experiments, facilitates a local power analysis. To this end, parametrized families of alternatives extending the so-called bivariate Konijn alternatives (Konijn, 1956)²³ are considered within which tests based on center-outward rank generalized symmetric covariances are shown to achieve rate-optimality (with root- n rate). Simulations provide strong evidence of the superiority of center-outward rank-based tests over, e.g., the traditional distance covariance test.

More recently, Deb et al. (2020) broadened the scope of this strand of literature to general kernel-based measures of dependence between variables \mathbf{X} and \mathbf{Y} with values in general topological spaces \mathcal{X} and \mathcal{Y} , respectively. The kernel measures of association η_K they introduce are not just a class of test statistics for the null hypothesis of vector independence: they also measure the strength of that dependence, in the sense that they range over $[0, 1]$, with value zero iff \mathbf{X} and \mathbf{Y} are independent and value one iff there exists some measurable function g such that $\mathbf{Y} = g(\mathbf{X})$, say (η is not necessarily symmetric in \mathbf{X} and \mathbf{Y}). The population version of η_K involves a kernel K (an adequately²⁴ symmetric, nonnegative definite function on $\mathcal{Y} \times \mathcal{Y}$). Its sample version $\hat{\eta}_n$ moreover relies on some adequate graph functional (such as a k -nearest neighbor or a minimum spanning tree graph).²⁵

As far as measure transportation-based methods are concerned (Section 8 of their paper), however, they are assuming that \mathbf{X} and \mathbf{Y} are real-valued random vectors with dimensions $d^{\mathbf{X}}$ and $d^{\mathbf{Y}}$, respectively. A rank-based version of kernel-based measures of dependence then is readily obtained by replacing, in $\hat{\eta}_n$, \mathbf{X} and \mathbf{Y} with their empirical vector ranks $\mathbf{F}_{\text{MK};\mathbf{X}}^{(n)}$

²³Konijn alternatives are classical in the context of testing for bivariate independence: see, e.g., Gieser and Randles (1997), Taskinen et al. (2005), Hallin and Paindaveine (2008a).

²⁴We are skipping details.

²⁵We are skipping details.

and $\mathbf{F}_{\text{MK};\mathbf{Y}}^{(n)}$. An attractive feature of the resulting rank-based statistics, which under the null hypothesis of independence between \mathbf{X} and \mathbf{Y} are distribution-free, is the Gaussian nature of their asymptotic null distributions, allowing for readily available critical values; this is in sharp contrast with the mixture of chi-squares obtained by Deb and Sen (2019) and Shi et al. (2021a, 2020).²⁶ Asymptotic normality is established via a Hájek asymptotic representation result—limited to identity (Spearman or Wilcoxon) score functions since \mathbf{X} and \mathbf{Y} are “simply” replaced with $\mathbf{F}_{\text{MK};\mathbf{X}}^{(n)}$ and $\mathbf{F}_{\text{MK};\mathbf{Y}}^{(n)}$. The paper concludes with a numerical comparison of the finite-sample performance of the proposed tests with the original distance covariance tests and the Hilbert–Schmidt independence tests (involving kernel-based measures of dependence) developed in the learning literature (see, e.g., Gretton et al. (2005a), Fukumizu et al. (2008)). Simulation results indicate that the methods proposed in the paper are competitive against dependencies under which small variations of \mathbf{X} induce large variations of \mathbf{Y} , and somewhat less competitive against smoother relationships. The power, however, largely depends on the graph structure considered for finite-sample implementation—e.g., the choice of k in the k -nearest neighbor graph approach. No comparison with the center-outward methods developed in Shi et al. (2020) is conducted; although scores (more general than Wilcoxon or Spearman) may improve performance significantly, final conclusions and consistency rates²⁵ are likely to remain unchanged.

Bibliographic note

The contributions by Ghosal and Sen and Deb and Sen on one hand, by Shi, Drton, and Han²⁷ on the other hand, were obtained independently and posted on arXiv almost contemporaneously (May 14, September 18, and September 22, respectively) in 2019. Although addressing the same problem, they (as well as the sections on rank-based statistics in Deb et al. (2020)) use different methods, showing little overlap and profitably complementing each other.

4.1.4 Goodness-of-fit and symmetry

Based on the techniques developed for testing vector independence, Ghosal and Sen (2019) and Deb and Sen (2019) also propose rank-based tests for the two-sample goodness-of-fit problem (see also Boeckel et al. (2018) and Hallin et al. (2021b)) and the null hypothesis of symmetry. The spirit of their approach is quite similar to that developed for testing vector independence, with (for Deb and Sen (2019)) tests based on *energy statistics* (Székely and Rizzo, 2013) rather than distance covariance; we refer to the papers for further details.

4.2 Multiple-output quantile regression

Shifting from rank-based to quantile-based inference, the main application of measure transportation in the literature is multiple-output quantile regression. Single-output quantile re-

²⁶Shi et al. (2021b), however, are reporting some suboptimality results on consistency rates.

²⁷Under the title “Distribution-free consistent independence tests via Hallin’s multivariate ranks.”

gression, as introduced in the seminal paper by Koenker and Bassett (1978), constitutes an ensemble of statistical techniques intended to estimate and draw inferences about conditional quantile functions. Considerably more informative than traditional mean regression, it has become a central tool in regression analysis—see Koenker et al. (2017) for an extensive recent survey of related topics. Desirable as it is, an adequate multiple-output version of the method has remained an open problem for almost half of a century—due, mainly, to the absence of an adequate concept of multivariate (conditional) quantile. Attempts have been made based on (conditional) halfspace depth (Hallin et al., 2010a, 2015) but halfspace depth contours and regions do not enjoy all the properties a quantile function is expected to satisfy—in particular, depth contours (the collection of points with given depth τ) are intrinsically convex and the probability of the enclosed region depends on the actual underlying distribution. The measure transportation concept of vector quantile is giving this problem an elegant and computationally feasible solution, developed in Carlier et al. (2016); see also Carlier et al. (2017, 2020).

Carlier et al. (2016) define the concept of conditional vector quantiles and use it to propose a model of (linear) vector quantile regression extending to the multiple-output case the traditional quantile regression model of Koenker and Bassett (1978). Let \mathbf{Y} , with values in \mathbb{R}^{d^Y} be a vector of dependent variables and denote by \mathbf{Z} , with values in \mathbb{R}^{d^Z} , a vector of covariates. Assume that (\mathbf{Y}, \mathbf{Z}) have joint distribution $P_{\mathbf{Y}\mathbf{Z}}$ and that \mathbf{Y} has a conditional (on $\mathbf{Z} = \mathbf{z}$) distribution $P_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}$ for all \mathbf{z} in the support of \mathbf{Z} . Let U denote some Lebesgue-absolutely continuous d^Y -dimensional reference distribution with convex support. A *vector quantile function of \mathbf{Y} conditional on \mathbf{Z}* is a mapping from $(\mathbf{u}, \mathbf{z}) \in \mathbb{R}^{d^Y} \times \mathbb{R}^{d^Z}$ to $\mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z}) \in \mathbb{R}^{d^Y}$ such that (i) for all $\mathbf{z} \in \mathbb{R}^{d^Z}$, $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z})$ is the gradient²⁸ of a convex function and (ii) $\mathbf{Y} = \mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{U}, \mathbf{Z})$ for some random vector \mathbf{U} with conditional distribution U given $\mathbf{Z} = \mathbf{z}$. Condition (i) implies that, for any \mathbf{z} , $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z})$ pushes U forward to $P_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}$; condition (ii) ensures the existence of a strong representation of \mathbf{Y} in terms of \mathbf{Z} and a latent random vector \mathbf{U} with distribution U . The theory of measure transportation and, more particularly, Mc Cann’s theorem then guarantee the existence and for-all- \mathbf{z} - U -a.s. unique vector quantile function for \mathbf{Y} conditional on \mathbf{Z} . If, moreover, for all \mathbf{z} in the support of \mathbf{Z} , $P_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}$ admits a density, then $\mathbf{u} \mapsto \mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z})$ admits an inverse²⁹ with the interpretation of a *vector rank mapping*. This extends to the conditional case the unconditional concepts of Monge-Kantorovich vector ranks and quantiles developed (subsequently) in Chernozhukov et al. (2017); a straightforward definition of Monge-Kantorovich regression depth and regression depth contour follows.

If $\mathbf{z} \mapsto \mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z})$ is regular enough, it characterizes the impact, on \mathbf{Y} ’s “quantile of order \mathbf{u} ,” of the values \mathbf{z} of the covariates—yielding a multiple-output quantile regression equation of \mathbf{Y} with respect to \mathbf{Z} . Assuming for all \mathbf{u} and, possibly, for some given d^X -dimensional function $\mathbf{X} := f(\mathbf{Z})$ of \mathbf{Z} , a linear form of this equation yields a *linear vector quantile regression*

²⁸As usual, this means “the Lebesgue-a.e.-gradient” of a convex function.

²⁹More precisely, for all \mathbf{z} , a mapping $\mathbf{y} \mapsto \mathbf{F}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}, \mathbf{z})$ such that $\mathbf{F}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Q}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{u}, \mathbf{z}), \mathbf{z}) = \mathbf{u}$ and $\mathbf{F}_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}, \mathbf{Z}) = \mathbf{U}$ U-a.s.

model

$$\mathbf{Y} = \mathbf{Q}_{\mathbf{Y}|\mathbf{X}}(\mathbf{U}, \mathbf{X}) = \beta'(\mathbf{U})\mathbf{X}$$

where $\mathbf{u} \mapsto \beta(\mathbf{u})$ is a $d^{\mathbf{X}} \times d^{\mathbf{Y}}$ matrix-valued function such that $\mathbf{u} \mapsto \beta'(\mathbf{u})\mathbf{x}$ is the gradient of a convex function and \mathbf{U} has, conditionally on \mathbf{X} , reference distribution U . A linear programming solution is proposed for the estimation of $\mathbf{u} \mapsto \beta(\mathbf{u})$, which embeds the vector quantile estimation algorithm proposed in Chernozhukov et al. (2017).

4.3 Nonlinear Principal Components (PC) and Independent Components (IC)

Principal Component (PC) and Independent Component (IC) methods are two major tools, part of daily practice in basically all domains of data analysis, statistical learning, and signal theory. Being based on linear transformations, they suffer, however, of severe limitations: principal components only account for linear dependencies, while independent components, which more ambitiously aim at turning a vector \mathbf{Z} of mutually interrelated observations into a vector \mathbf{U} of independent variables, fail to do so unless ad hoc assumptions are made on the data-generating process.

By abandoning the restriction to linear transformations, measure transportation with reference distribution U factorizing into a product of univariate marginals, brings, under very mild assumptions on \mathbf{Z} , a natural solution to the problem. Two reference distributions of interest, in that context, are the uniform $U_{[0,1]^d}$ over the unit cube $[0, 1]^d$ or any of its rotations $\mathbf{O}[0, 1]^d$ (\mathbf{O} an arbitrary orthogonal matrix) and the spherical $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ standard normal. Adopting the terminology of Section 3.3.2, for any orthogonal $d \times d$ matrix \mathbf{O} , the vector rank mapping $\mathbf{F}_{\text{MK}}^{\mathbf{O}}$ of $\mathbf{O}\mathbf{Z}$ with reference $U = U_{[0,1]^d}$ and $\mathbf{O}\mathbf{F}_{\text{MK}}^{\mathcal{N}}$, where $\mathbf{F}_{\text{MK}}^{\mathcal{N}}$ denotes the vector rank mapping of \mathbf{Z} with reference $U = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ define nonlinear independent component transformations of \mathbf{Z} .

That idea of a nonlinear independent component analysis based on measure transportation has been developed by Gunsilius and Schennach (2019), along with an entropy-based selection of \mathbf{O} , yielding a d -tuple of independent components ordered by decreasing information content and an entropy decomposition that parallels the classical decomposition of \mathbf{Z} 's total variance into the sum the variances of its principal components.³⁰ That entropy decomposition naturally suggests a nonlinear dimension reduction procedure.³¹ When $\mathbf{F}_{\text{MK}}^{\mathbf{O}}$ and $\mathbf{F}_{\text{MK}}^{\mathcal{N}}$ are homeomorphisms, they admit continuous inverses $\mathbf{Q}_{\text{MK}}^{\mathbf{O}}$ and $\mathbf{Q}_{\text{MK}}^{\mathcal{N}}$ mapping the nonlinear independent components to a system of curves providing a graphical representation that greatly helps their intuitive interpretation. Gunsilius and Schennach (2019) also propose an estimation method which involves the analytical properties of the vector rank map and a kernel estimation of \mathbf{Z} 's

³⁰When \mathbf{Z} itself is Gaussian, the method coincides with the standard linear principal component decomposition.

³¹A number of nonlinear dimension reduction techniques have been proposed, mainly in the machine learning literature, but seldom produce mutually independent components. Moreover, their consistency properties, in general, are not clear, and many of them are lacking a population version. Section 3 of Gunsilius and Schennach (2019) provides an overview of that literature which we do not reproduce here.

density. This estimation step perhaps could be revised in light, e.g., of the results by Deb and Sen (2019) (which probably were not available to the authors when their paper was written).

5 CONCLUSIONS AND SOME DIRECTIONS FOR FUTURE RESEARCH

Measure transportation ideas and techniques, in a very short span of time, (i) have been shown to provide, unlike previous attempts, theoretically adequate multivariate extensions of the dual concepts of distribution and quantile functions (in population), of ranks and quantiles (in the sample) and (ii) have demonstrated their applicability in a variety of contexts—distribution-free hypothesis testing, multiple-output quantile regression, nonlinear principal and independent component analysis, nonlinear dimension reduction, etc. Many questions remain open, however:

(a) several issues remain to be studied about the concepts themselves: how in finite samples should we choose the factorization of n into $n_R n_S + n_0$ in the definition of center-outward ranks and signs? should we consider cross-validation? random grids?

(b) can powerful goodness-of-fit tests be based, e.g. on Kolmogorov-Smirnov or Cramér-von Mises distances between center-outward distribution functions, on the model of Ghosal and Sen (2019)?

(c) turning to quantiles, what are the properties of $\mathbf{Q}_{\pm}^{(n)}(\mathbf{0})$ (for $n_0 \neq 0$) as a multivariate median? can we construct multivariate median or sign tests? can we, combining ideas from Carlier et al. (2016) and Hallin et al. (2015), perform local bilinear multiple-output center-outward quantile regression? construct multivariate growthcharts?

(d) 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 notions such as Lorenz curves, average values at risk, or expected shortfall, on the model of Gushchin and Borzykh (2017) and Beirlant et al. (2020)?

(e) what happens in high dimension ($d \rightarrow \infty$)? in functional spaces? on spheres (directional data)? on other Riemannian manifolds?

(f) finally, new empirical distribution and quantile functions are calling for an in-depth study of the corresponding empirical processes with further results such as Donsker and iterated logarithm theorems or Bahadur representations.

References

- Bakirov, N. K., Rizzo, M. L., and Székely, G. J. (2006). A multivariate nonparametric test of independence. *J. Multivariate Anal.*, 97:1742–56.
- Basu, D. (1959). The family of ancillary statistics. *Sankhyā Ser. A*, 21:247–56.
- Beirlant, J., Buitendag, S., del Barrio, E., Hallin, M., and Kamper, F. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math. Econom.*, 95:79–100.

- Belloni, A. and Winkler, R. (2011). On multivariate quantiles under partial orders. *Ann. Statist.*, 39:1125–79.
- Boeckel, M., Spokoiny, V., and Suvorikova, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing. Available at arXiv:1809.04090.
- Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417.
- Carlier, G., Chernozhukov, V., De Bie, G., and Galichon, A. (2020). Vector quantile regression and optimal transport, from theory to numerics. *Empirical Economics*, 0:0–0.
- Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression: an optimal transport approach. *Ann. Statist.*, 44:1165–92.
- Carlier, G., Chernozhukov, V., and Galichon, A. (2017). Vector quantile regression beyond the specified case. *J. Multivariate Anal.*, 161:96–102.
- Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, 91:862–72.
- Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs. *Ann. Statist.*, 45:223–56.
- Chernozhukov, V., Galichon, A., Henry, M., and Pass, B. (2020). Identification of hedonic equilibrium and nonseparable simultaneous equations. *J. Political Econ.*, DOI = 10.1086/712447.
- Cuesta-Albertos, J. and Matrán, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.*, 17:1264–76.
- de Valk, C. and Segers, J. (2019). Tails of optimal transport plans for regularly varying probability measures. Available at arXiv:1811.12061.
- Deb, N., Ghosal, P., and Sen, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. Available at arXiv:2010.01768.
- Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. Available at arXiv:1909.08733.
- del Barrio, E., Cuesta-Albertos, J., Hallin, M., and Matrán, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions. Available at arXiv:1806.01238.
- del Barrio, E., González-Sanz, A., and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.*, 180:104671, 13.

- Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration. In Chen, A. S. and Travaglini, G., Eds., *Panorama of Discrepancy Theory*, pages 539–620. Springer, New York.
- Ekeland, I., Galichon, A., and Henry, M. (2012). Comonotonic measures of multivariate risks. *Math. Finance*, 22:109–32.
- Faugeras, O. and Rüschendorf, L. (2017). Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles. *Mathematica Applicanda*, 45:21–63.
- Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.*, 177 :413–21.
- Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. (2008). Kernel measures of conditional dependence. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S. T., editors, *Advances in Neural Information Processing Systems 20*, pages 673–80. Curran, Red Hook, NY.
- Galichon, A. and Henry, M. (2012). Dual theory of choice with multivariate risks. *J. Econ. Th.*, 147:1501–16.
- Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.*, 47:551–79.
- Ghosal, P. and Sen, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. Available at arXiv:1905.05340.
- Gieser, P. W. and Randles, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.*, 92:561–67.
- Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005a). Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory*, pages 63–77. Springer, Berlin, Germany.
- Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S., editors, *Advances in Neural Information Processing Systems 20*, pages 984–91. Curran, Red Hook, NY.
- Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b). Kernel methods for measuring independence. *J. Mach. Learn. Res.*, 6:2075–2129.
- Gretton, A., Smola, A., Bousquet, O., Herbrich, R., Belitski, A., Augath, M., Murayama, Y., Pauls, J., Schölkopf, B., and Logothetis, N. (2005c). Kernel constrained covariance for dependence measurement. In Cowell, R. G. and Ghahramani, Z., editors, *AISTATS05*, pages 112–119. Society for Artificial Intelligence and Statistics.

- Gunsilius, F. and Schennach, S. M. (2019). Independent nonlinear component analysis. CEMMAP Working Paper CWP46/19, London.
- Gushchin, A. and Borzykh, D. (2017). Integrated quantile functions: properties and applications. *Mod. Stoch. Theory Appl.*, 4:285–314.
- Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
- Hallin, M. (2017). On distribution and quantile functions, ranks and signs in \mathbb{R}^d : a measure transportation approach. Available at ideas.repec.org/p/eca/wpaper/2013-258262.html.
- Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2021a). Distribution and quantile functions, ranks, and signs in \mathbb{R}^d : a measure transportation approach. *Ann. Statist.* (in press).
- Hallin, M., Hlubinka, D., and Hudecová, Š. (2020a). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. Available at [arXiv:2007.15496](https://arxiv.org/abs/2007.15496).
- Hallin, M., Ingenbleek, J.-F., and Puri, M. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.*, 30:34–71.
- Hallin, M., La Vecchia, D., and Liu, H. (2020+b). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.*, pages 1–14. (in press).
- Hallin, M., La Vecchia, D., and Liu, H. (2020+c). Rank-based testing for semiparametric var models: a measure transportation approach. Available at [arXiv:2011.06062](https://arxiv.org/abs/2011.06062).
- Hallin, M., Lu, Z., Paindaveine, D., and Šiman, M. (2015). Local bilinear multiple-output quantile regression. *Bernoulli*, 21:1435–66.
- Hallin, M., Mordant, G., and Segers, J. (2021b). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electronic Journal of Statistics*, 15:1–44.
- Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, 30:1103–33.
- Hallin, M. and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.*, 32:2642–78.
- Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape: I optimal rank-based tests for sphericity. *Ann. Statist.*, 34:2707–56.
- Hallin, M. and Paindaveine, D. (2008a). Chernoff-Savage and Hodges-Lehmann results for Wilks’ test of multivariate independence. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in honor of Professor Pranab K. Sen*, pages 184–96. I.M.S., Beachwood, OH.

- Hallin, M. and Paindaveine, D. (2008b). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.*, 36:1261–98.
- Hallin, M., Paindaveine, D., and Šiman, M. (2010a). Multivariate quantiles and multiple-output regression quantiles: from L^1 optimization to halfspace depth (with discussion and rejoinder). *Ann. Statist.*, 38:635–69.
- Hallin, M., Paindaveine, D., and Verdebout, T. (2010b). Optimal rank-based testing for principal components. *Ann. Statist.*, 38:3245–99.
- Hallin, M., Paindaveine, D., and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli*, 19:2524–56.
- Hallin, M. and Puri, M. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.*, 50:175–237.
- Hallin, M. and Werker, B. (1998). Optimal testing for semiparametric autoregressive models: from Gaussian Lagrange multipliers to regression rank scores and adaptive tests. In Subir Ghosh, Ed., *Asymptotics, Nonparametrics, and Time Series*, pages 295–358. Marcel Dekker, New York.
- Hallin, M. and Werker, B. (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli*, 9:137–65.
- Hamel, A. and Kostner, D. (2018). Cone distribution functions and quantiles for multivariate random variables. *J. Multivariate Anal.*, 167:97–113.
- Hodges, J. (1955). A bivariate sign test. *Ann. Math. Statist.*, 26:523–27.
- Judd, K. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, Mass.
- Kim, I., Balakrishnan, S., and Wasserman, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.*, 48:3417–41.
- Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica*, 46:33–50.
- Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. CRC Press.
- Konijn, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.*, 27:300–23.
- Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer, New York.
- Liu, R. (1992). Data depth and multivariate rank tests. In Y. Dodge, Ed., *L^1 Statistics and Related Methods*, pages 279–94. North-Holland, Amsterdam.

- Liu, R. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88:257–60.
- López-Pintado, S. and Romo, J. (2012). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, 104:718–34.
- Marden, J. (1999). Multivariate rank tests. In S. Ghosh, Ed., *Design of Experiments and Survey Sampling*, pages 401–32. Marcel Dekker, New York.
- McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–23.
- Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, PA.
- Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, 26:319–43.
- Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks*. Springer, New York.
- Oja, H. and Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes. *J. Statist. Plann. Inference*, 135:300–23.
- Oja, H. and Randles, R. H. (2004). Multivariate nonparametric tests. *Statist. Sci.*, 19:598–605.
- Panaretos, V. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, 6:405–31.
- Peyré, G. and Cuturi, M. (2019). Computational optimal transport with applications to Data Science. *Found. Trends Machine Learning*, 11:355–607.
- Puri, M. and Sen, P. (1985). *Nonparametric Methods in General Linear Models*. Wiley, New York.
- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley, New York.
- Rachev, S. T. and Rüschendorf, L. (1998). *Mass Transportation Problems*. Springer, New York.
- Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.*, 84:1045–50.
- Rüschendorf, L. and Rachev, S. T. (1990). A characterization of random variables with minimum L^2 -distance. *J. Multivariate Anal.*, 32:48–54.
- Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

- Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, 41:2263–91.
- Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerl.*, 56:214–32.
- Shi, H., Drton, M., and Han, F. (2021a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.*, 0(0):1–16.
- Shi, H., Drton, M., and Han, F. (2021b). On the power of Chatterjee rank correlation. Available at arXiv:2008.11619.
- Shi, H., Hallin, M., Drton, M., and Han, F. (2020). Rate-optimality of consistent distribution-free tests of independence based on center-outward ranks and signs. Available at arXiv:2007.02186.
- Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.*, 3:1236–65.
- Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: a class of statistics based on distances. *J. Statist. Plann. Inference*, 143:1249–72.
- Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35:2769–94.
- Taskinen, S., Oja, H., and Randles, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.*, 100:916–25.
- Um, Y. and Randles, R. H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist. Sinica*, 8:801–12.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer, New York.
- Villani, C. (2003). *Topics in Optimal Transportation*. Amer. Math. Soc., Providence, RI.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer-Verlag, Berlin, Germany.
- Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 105:547–62.
- Yosida, K. (1964). *Functional Analysis*. Springer, New York.
- Zhu, L., Xu, K., Li, R., and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika*, 104:829–43.
- Zuo, Y. (2018). On general notions of depth for regression. *arXiv:1805.02046*.
- Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, 34:2879–96.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, 28:461–82.