On Multivariate Distribution and Quantile Functions, Ranks and Signs

a measure transportation approach

Marc Hallin
ECARES and Département de Mathématique, Université libre de Bruxelles

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