Optimal Transport with Application to Quantile Regression

In dimension one, the quantile is the inverse of the cumulative distribution function. Quantiles are extremely useful objects because they fully characterize the distribution of an outcome, and they allow to provide directly a number of statistics of interest such as the median, the extremes, the deciles, etc. They allow to express maximal dependence between two random variables (comonotonicity) and are used in decision theory (rank-dependent expected utility); in finance (value-at-risk and Tail VaR); in microeconomic theory (efficient risksharing); in macroeconomics (inequality); in biometric (growth charts) among other disciplines. Quantile regression, pioneered by Roger Koenker, allows to model the dependence of a outcome with respect to a set of explanatory variables in a very flexible way, and has become extremely popular in econometrics.
The classical definition of quantiles based on the cumulative distribution function, however, does not lend itself well to a multivariate extension, and given the number of applications of the notion of quantiles, many authors have suggested various proposals.
We have proposed a novel definition of multivariate quantiles called "Vector quantiles" based on optimal transport. The idea is that instead of viewing the quantile map as the inverse of the cumulative distribution function, it is more fruitful to view it as the map that rescales a distribution of interest to the uniform distribution over [0,1] in the least possible distortive way, in the sense that the average squared distance between an outcome and its preimage by the map should be minimized. While equivalent to the classical definition in the univariate case, this definition lends itself to a natural multivariate generalization using Monge-Kantorovich theory. We thus define the multivariate quantile map as the one that attains the L2-Wasserstein distance, which is known in the optimal transport literature as Brenier's map. With Carlier and Santambrogio, we gave a precise connection between vector quantiles and a celebrated earlier proposal by Rosenblatt. We have shown that many of the desirable properties and uses of univariate quantiles extend to the multivariate case if using our definition. With Ekeland and Henry, we have shown that this notion is a multivariate analog of the notion of Tail VaR used in finance. With Henry, we have shown that it allows to construct an extension of Yaari's rank-dependent utility in decision theory, and have provided an axiomatization for it. With Carlier and Dana, we have shown that this notion extends Landsberger and Meilijson's celebrated characterization of efficient risksharing arrangements. With Charpentier and Henry, we have shown the connection with Machina's theory of local utility. With Carlier and Chernozhukov, we have shown that Koenker's quantile regression can be naturally extended to the case when the dependent variable is multivariate if one adopts our definition of multivariate quantile. With Chernozhukov, Hallin and Henry we have defined empirical vector quantiles and have studied their consistency.
A number of challenges remains. Among these, an empirical process theory for multivariate quantiles (extending the univariate theory of the empirical quantile process) is the obvious next step, although this is highly non-trivial. Invariance issues are also interesting and challenging questions. Finally, while the link with other multivariate quantiles (such as Rosenblatt's) is by now well understood, the link with other related concepts, such as Tukey's halfspace depth, remains to be explored.

Based on joint work with Guillaume Carlier, Arthur Charpentier, Victor Chernozhukov, Rose-Anne Dana, Ivar Ekeland, Marc Hallin, Marc Henry, and Filippo Santambrogio.

REFERENCES:

Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio (2010). From Knothe's transport to Brenier's map. *SIAM Journal on Mathematical Analysis* 41, Issue 6, pp. 2554-2576. Available here.

Ivar Ekeland, Alfred Galichon, and Marc Henry (2012). Comonotonic measures of multivariate risks. *Mathematical Finance* 22 (1), pp. 109-132. Available here.

Alfred Galichon and Marc Henry (2012). Dual theory of choice with multivariate risks. *Journal of Economic Theory* 147(4), pp. 1501–1516. Available here.

Guillaume Carlier, Rose-Anne Dana, and Alfred Galichon (2012). Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory* 147(1), pp. 207–229. Available here.

Arthur Charpentier, Alfred Galichon, and Marc Henry (2016). Local utility and risk aversion. *Mathematics of Operations Research* 41(2), pp. 466—476.

Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon (2016). Vector quantile regression: an optimal transport approach.*Annals of Statistics* 44 (3), pp. 1165–1192. Available here. Software available here.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry (2016). Monge-Kantorovich Depth, Quantiles, Ranks and Signs. *Annals of Statistics.*Available here.

Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon (2017). Vector quantile regression beyond correct specification. *Journal of Multivariate Analysis.* Available here.