Bridging centrality and extremity: refining sample data depth using extreme value theory

Prof. Dr. John H.J. Einmahl
(Tilburg University, The Netherlands)

June 15, 2011

A data depth is a measure of centrality of a point with respect to a given probability distribution or data cloud. It provides a natural center-outward ordering of multivariate data points, and yields a new nonparametric multivariate analysis scheme. Thus far, data depth approaches to multivariate analysis have had many successes. In particular, the approaches derived from geometric depths (e.g. half-space depth) are especially useful since they capture the true probabilistic geometry underlying the data. However, the sample geometric depth is defined to be zero once it is outside the convex hull of the data set. This property has limited the utility of depth approach in applications, such as in classification problems and control charts with extremely small false alarm rates. To address this issue, we propose to apply extreme value theory to refine the empirical half-space depth in the tails of the data set. This provides an interesting linkage between data depth, which is useful for inference on centrality, and extreme value theory, which is useful for inference on extremes. The refined estimator of data depth broadens greatly the applicability of data depth. This is work in progress and therefore, after a detailed introduction, we will focus on the construction of the estimator and the theory, starting with the one-dimensional case. Joint work with Jun Li (University of California, Riverside) and Regina Y. Liu (Rutgers University).