

Data Depth and Nonparametric Multivariate Statistics: Spacings, Ordering and Beyond

Regina LIU, Jun LI and Juan CUESTA

Rutgers University, UC Riverside and University of Cantabria, Spain

Abstract: There has been considerable new interest in nonparametric multivariate statistics due to the development of data depth and its induced center-outward ordering (or ranking) of multivariate data. We highlight some of the recent advances. In particular, we introduce and develop multivariate spacings using the order statistics derived from data depth. Specifically, the spacing between two consecutive order statistics is the region which bridges the two order statistics, in the sense that the region contains all the points whose depth values fall between the depth values of the two consecutive order statistics. These multivariate spacings can be viewed as a data-driven realization of the so-called “statistically equivalent blocks”. These spacings assume a form of center-outward layers of “shells” (“rings” in the two-dimensional case), for which the shapes of the shells follow closely the underlying probabilistic geometry. We discuss the properties and applications of these spacings. For example, we use the spacings to construct tolerance regions. The construction is nonparametric and completely data driven, and the resulting tolerance region reflects the true geometry of the underlying distribution. This is different from the existing approaches which require that the shape of the tolerance region be specified in advance. Finally, we also discuss multivariate goodness-of-fit tests based on the proposed spacings.