On nonparametric depth based classification of functional observations

Sami Helander[1], Stanislav Nagy[2], Germain Van Bever[3], Lauri Viitasaari[4] and Pauliina Ilmonen[1]

[1] *Aalto University School of Science, Finland, sami.helander@aalto.fi, pauliina.ilmonen@aalto.fi*
[2] *Charles University, Czech Republic, nagy@karlin.mff.cuni.cz*
[3] *Université libre de Bruxelles (ULB), Belgium, gvbever@ulb.ac.be*
[4] *University of Helsinki, Finland, lauri.viitasaari@iki.fi*

Keywords: functional data analysis, statistical depth, classification.

Lack of storage capacity and computing power limit us very little these days. We have been able to go from multivariate data to very high dimensional data. One approach used when dealing with such high dimensional data is to assume that the observed units are random functions (from some generating process) instead of random vectors.

The concept of statistical depth was originally introduced as a way to provide a nonparametric center-outward ordering from a depth-based multivariate median. Several different depth functions for functional observations have been presented in the literature. Most of these approaches, however, are based on assessing the location of the function as a measure of typicality. As a result, they are missing some important features inherent to functional data such as variation in shape. Another problem in assessing typicality of functional observations and in classifying functional observations is that we often observe only part of the function. One may overcome this problem by extrapolating and interpolating i.e. by adding the missing parts. However, doing that, at least implicitly, requires model assumptions.

We discuss assessing typicality of functional observations. Moreover, we provide a new classification method that is based on j-th order k-th moment integrated depths. For $j = 1$ and $k = 1$ this is equal to applying the mean halfspace depth of a functional value with respect to the corresponding univariate marginal distribution. When j is larger than 1, the method is not based on comparing location only but considers shape of the function as well. Moreover, the method can be applied to partially observed functions without extrapolation or interpolation. Theoretical properties of the new approach are explored and several real data examples are presented to demonstrate its excellent classification performance.