Discussion of "Multivariate functional outlier detection" by Mia Hubert, Peter Rousseeuw and Pieter Segaert

Karl Mosler

Mia Hubert, Peter Rousseeuw and Pieter Segaert (subsequently HRS) are to be praised for having developed a visual and well applicable methodology of treating outliers in functional data. Their principal achievements include a systematics of functional outliers, novel visualizations and a measure of outlyingness (bagdistance), which nicely combines Euclidean distance with location depth. Most important, the authors offer a consequent multivariate view on the functions, either assuming genuinely multidimensional functions or considering multiple aspects of such functions. Several real-data examples illuminate the capacity and the possible output of the new procedures. Also, an R-package of the methodology is announced, which will be greatly welcomed.

In the following remarks, I shall first try to put the new methodology into a broader context of outlier search. Then some specifics of functional data are discussed, which call for particular treatments. Third a refined approach is sketched that considers integrals over subintervals. Finally, I will address the use of different functional depth notions and the computational problems arising with them.

1 The broad picture

The task of outlier detection is to distinguish 'irregular' observations from 'regular' ones. The latter observations are generated by a 'regular process', while the others arise from one (or several) 'contaminators'. As in every statistical inquiry, the first question to be posed is: What do we *a priori* know about the regular generating process? Which model can be reasonably chosen for it? Of course, the answer depends primarily on the particular application at hand.

K. Mosler
University of Cologne, Albertus-Magnus-Platz, 50923 Cologne
Tel.: +49-221-470-2982
Fax: ++49-221-470-5084
E-mail: mosler@statistik.uni-koeln.de

With functional data, in most applications some prior knowledge is given how it has been produced or how it should look by its nature. It may result from the problem setting, e.g., that the functions are monotonic (or bounded to an interval, or periodic), or that they fit to paths of some ARMA or GARCH process, or similar.

But, other than standard statistical problems, the problem of outlier detection needs additional prior considerations. A second question must be answered in some way: Which sort of outlying observations do we want to detect and eventually eliminate? This depends to some part on the application at hand (Which data contaminations have been earlier occurred or are likely due to the procedure of measurement?), but to a larger part on the goal of the eventual statistical inquiry. E.g., if inference shall be made about a location parameter, shift outliers will be relevant, rather than dispersion outliers. Specifically for functional data, many more possible contaminations have to be thought of than for univariate or multivariate data. We must decide, at least implicitly, which of the numerous possible deviations actually turn a function into an outlier. As an extreme case, consider the problem of outliers in electrocardiogram (ECG) data, which is mentioned in the paper. Such data are searched for pathological deviations, which correspond to certain heart diseases. In this application, extensive medical information is available on pathological shapes, so that the outlier problem rather becomes a problem of supervised classification.

HRS suggest a practical approach to answer Question 2 by offering a partial taxonomy of outlyingness in functional data: They first distinguish an isolated (= peak) outlier from a persistent one. Among the latter they discern shift outliers and amplitude outliers. The rest is called outliers of 'shape', which, among others, include outliers regarding slope or phase. Several of the mentioned categories come with a class of data transformations, by which an outlier is turned into a regular observation: a shift outlier by a baseline correction, an amplitude outlier by a rescaling, a phase outlier by a warping transformation.

A third question to be answered regards the proper statistical procedure of outlier detection. Which statistics shall be used that identifies and eliminates the undesirable outliers from the given data? Following their taxonomy of different aspects of outlyingness, HRS provide diagnostic functions so that outlyingness in a certain aspect is detected through locational outlyingness of the respective diagnostic function. Further they offer diagnostic plots, by which such outliers are visualized. For the categories mentioned, this approach proves to be clever and practical, which is demonstrated by examples.

More general, the HRS approach says: Construct auxiliary diagnostic functions, like derivatives and warping functions, whose large deviations correspond to outlyingness in the given application. Augment and partially substitute the functional observations by the diagnostic functions. If k is the dimension of the augmented functions, construct a 50% depth-trimmed region (the bag) in $\mathbb{R}^k$ and calculate the bagdistances of all augmented observations. In other words, the HRS approach adds functional features to the observations so that the outliers sought for are detected as *locational outliers of the augmented data.*

Of course, the 'taxonomy' of HRS is not complete and the approach not universally applicable. However, their basic idea, to substitute or enhance the given functional observations by other functions that indicate a specific sort of outlyingness seems to have a promising range of applications.

2 Features of functional data

Functional data can be regarded under many points of view and described by a multitude of features. Specifically, functions depend on an argument, usually time or a function of time, and the development of a function over time has some meaning. Therefore, in comparing functions not only their levels may be taken into account but also their increasing behavior. For example, in the classic Berkeley data the growth curves of girls and boys are most easily distinguished by looking at the average slope of the curves above the age of 10; see Mosler & Mozharovskyi (2014). Most existing procedures for functional data analysis largely disregard this aspect by considering levels of functions only and aggregating this information symmetrically in the time parameter. This means that an arbitrary permutation of the time variable will not affect the result: the procedure is rearrangement invariant. A simple remedy is to include derivatives of the data into the analysis as HRS do, too. Specifically, slope outliers are identified by enhancing the observations with their first derivatives and searching them for local locational outliers. Another remedy consists in fitting a robust time series model to the data and doing inference about its parameters.

Functional data may also show shifts or other variations in their time argument. If such deviations are relevant in the eventual statistical inquiry, HRS propose to detect them by calculating standard warping functions and checking them for larger differences, that is, location outliers.

3 Local vs. global outlyingness: Considering subintervals

In their 'taxonomy', HRS distinguish two extreme cases of functional outliers, isolated and persistent ones: A deviation at any single time point (or very short interval) resp. a deviation over the whole interval (or a large part of it). Depending on the application, local deviations - as opposed to average deviations over the whole interval - may be expected and their identification be seen important. For example, deviations in level or phase can be local or global, while deviations in slope are local.

I think it is a key aspect of outlyingness and to be judged from the application at hand (see Question 2), up to which degree local deviations are regarded as outliers. This degree may be specified by some minimum length of subinterval over which a deviation should extend in order to constitute an outlier. To detect such local outliers, we split the original time interval into small enough subintervals and take averages over each of these intervals. (Taking

averages has the advantage that the original functional data can be used and no interpolation is needed: possible peaks are more or less evened. Also, if the data is given at discrete time points, these time points need not be common for all functional observations.) Then, a function is identified as outlying if it is outlying in at least one of the subintervals, which can be decided by the diagnostic functions and plots introduced in HRS (cf. e.g. their Figure 2).

To account for a possibly different importance of time regions HRS introduce a weight function in their multivariate functional depth (Equation (2)). However, such weights are difficult to obtain. Rather, as in Mosler & Mozharovskyi (2014), I suggest that the whole interval be divided into a number of subintervals of equal length, and the functions be averaged over the subintervals. If the number of subintervals is m, we obtain a problem of outlier detection in $\mathbb{R}^m$, which may be solved by employing a properly chosen multivariate data depth. The number of (equal length) subintervals has to be determined by evaluating some quality index.

4 Depth statistics for functional data

The bagdistance[1] of HRS appears to be a felicitous idea to quantify deviations from a center in a robust way. Here, the Tukey depth is only used to determine a center and a set-valued statistic (the bag) that, among others, reflects dispersion. The latter is employed for scaling the Euclidean distance of an observation from the center.

A more direct application of depth statistics is the following: Construct a functional data depth that is specific for the kind of sought outliers and identify those observations as outliers whose depth falls below some level. Many notions of depths for functional data have been introduced in the literature, and some have been applied to the outlier detection problem in this way; see e.g. the references given by HRS. In principle, each such depth notion will identify a different sort of outliers. If the depth is applied to higher-dimensional objects (like discretizations of the observed functions), only a few depth notions are feasible: They must not vanish outside the convex hull of the data and should be computable in reasonable time. Candidates are the Mahalanobis depth, the spatial depth, and - with some reservation - the projection depth. Note that the projection depth is very sensitive to small changes in direction. As it is piece-wise linear, the projection depth attains its maximum at the edges of direction cones of constant linearity, a randomly chosen direction yields the exact depth value with probability zero. Consequently, it has to be evaluated in a huge number of directions (which moreover should increase exponentially with dimension k), each of which involves the calculation of the median and MAD of a univariate distribution; see Mosler & Mozharovskyi (2014).

---

[1] The notion of bagdistance defined by HRS should not be mixed up with the notion of bag distance in data matching, that measures the similarity between strings; see Bartolini *et al.* (2002).

When it comes to computing, the HRS bagdistance needs the calculation of a Tukey-trimmed region (the bag). There exist recent efficient algorithms which calculate an exact Tukey region for dimensions k up to 9 (Liu *et al.*, 2015). As an alternative approach, HRS suggest the use of the *skew-adjusted projection depth* (SPD), which is approximately calculated as the infimum of univariate depths over a finite number of directions. However, I guess that the number of directions needed for a satisfactory approximation of the SPD is as large as that needed for the projection depth.

Finally, the functional depth approach extends to situations, where more than one function is typical for the regular process. If several (types of) functions are seen as typical, multiple centers can be considered. If the regular process is multimodal, it may be modelled by a location mixture. Then localized depth notions may be employed as e.g. in Sguera *et al.* (2015).

References

BARTOLINI, I., CIACCIA, P. AND PATELLA, M. (2002) String matching with metric trees using an approximate distance. In: String processing and information retrieval. Springer. pp 271–283.

LIU, X., MOSLER, K., AND MOZHAROVSKYI, P. (2015). Fast computation of Tukey trimmed regions in dimension $p > 2$. *Submitted.* `arXiv:1412.5122`.

MOSLER, K. AND MOZHAROVSKYI, P. (2014). Fast DD-classification of functional data. *Submitted.* `arXiv:1403.1158`.

SGUERA, C., GALEANO, P., AND LILLO, R.(2015) Functional outlier detection with a local spatial depth. `arXiv:1501.01859`.