Outlier Detection for Functional Data Using Principal Components

Matías Salibián-Barrera*
University of British Columbia, Vancouver, Canada – matias@stat.ubc.ca

Graciela Boente Boente
Universidad de Buenos Aires, Buenos Aires, Argentina – gboente@dm.uba.ar

Principal components analysis is a widely used technique that provides an optimal lower-dimensional approximation to multivariate observations in mean square error retaining as much information as possible. In the functional case, a new characterization of elliptical distributions on separable Hilbert spaces allows us to obtain an equivalent stochastic optimality property for the principal component subspaces of random elements on separable Hilbert spaces. This property holds even when second moments do not exist. Furthermore, these lower-dimensional approximations can be very useful in identifying potential outliers among high-dimensional or functional observations.

In this talk, we discuss the problem of estimating these finite-dimensional approximating linear subspaces robustly. The new class of robust estimators for principal directions is consistent for elliptical random vectors, and Fisher-consistent for elliptically distributed random elements on arbitrary Hilbert spaces. We illustrate our method on two real functional data sets, where the robust estimator is able to discover atypical observations in the data that would have been missed otherwise. Through a simulation study, we also study its performance when used to detect outlying observations.

This talk is the result of recent collaborations with Prof. David Tyler (Rutgers University).

Keywords: Functional Principal Components; Robustness; Functional Outliers; S-estimators.