Functional Data Smoothing Methods and Their Applications

Songqiao Huang
Department of Statistics
University of South Carolina, Columbia

SUMMARY: Functional data analysis was first introduced by Ramsay in 1982, who proposed viewing sequences of correlated time series as individual observations coming from intrinsic functional forms. Since then, many methods have been proposed to study the true underlying functions of the curves and the patterns among them, including functional applications on groups of curves. In functional data analysis, data smoothing usually serves as a preliminary but vital step, since the smoothness properties of the fitted curves directly affect any potential subsequent statistical analysis. Currently, many popular functional data smoothing methods have their own limitations: They are usually only suited to fit curves of specific shapes, or they require one to wisely choose tuning parameter values in order to obtain desirable performance. Motivated by these facts, in this dissertation, we explore innovative functional data smoothing methods that can be more easily adapted to functional curves of various shapes and are more flexible to use in practice. We first give a brief introduction to the history and recent research of functional data analysis and its applications. Then in Chapter 2, we propose a Bayesian functional data fitting method that uses transformed basis functions obtained via a "domain-warping" process based on the existing B-spline functions. Compared to many currently popular smoothing basis systems such as the B-spline basis, the Fourier basis, the Wavelet basis, etc., our transformed spline basis often achieves a better fit for the functional data as measured by smoothness and accuracy, while maintaining small to moderate model size. In Chapter 3, we investigate the importance of functional data smoothing methods on functional clustering and functional regression. We show via simulation and real data studies that our method proposed in Chapter 2 usually provides competitive improvements in the statistical

analysis results, compared with other popular functional data fitting methods. Acknowledging that the method proposed in Chapter 2 is suitable for fitting curves one by one with data that are observed over a fine grid of time points, in Chapter 4 we propose a method that extends the idea in order to fit a group of sparsely observed curves simultaneously. Our simulation studies show that our method produces very good estimates of the true structure of both the group mean curve and the individual effect curve for each observation.