

# **Dissimilarity for functional data clustering based on smoothing parameter commutation**

ShengLi Tzeng

Department of Applied Mathematics, National Sun Yat-sen University, Taiwan

Many studies measure the same type of information on the same subject at multiple time points, which are known as functional data. Patients' medical records have many valuable functional data, e.g., dose prescriptions over time. One way to turn such longitudinal collected data into clinically useful information is through cluster analysis, helping physicians learn more about metabolic change, disease evolution, etc. Various clusters reflect heterogeneity in patients' characteristics or effectiveness of therapy. Clustering such type of data, however, encounters two main challenges. First, not all patients were prescribed at regular time points. Hence missing values seems ubiquitous if one tries to align records at distinct time points. Second, a few outliers may heavily influence the estimation for within and/or between variations of clusters. This talk addresses a novel and easy-to-implement pairwise dissimilarity measure, which also serves as a screen tool to detect potential outliers. The method handles data observed at regular or irregular time points in the same way. The dissimilarity between subjects uses varying curve estimates with pairwise commutation of smoothing parameters. It takes into account the estimation uncertainty and, as a basis of outlier detection, it is not strongly affected by outliers. The effectiveness of the new proposal is shown by simulations and a real application to methadone dosage maintenance levels.

Keywords: clustering, functional data, smoothing splines, dissimilarity, outliers