

Analysis of Distribution Valued Data using Techniques of FDA

Masahiro Mizuta*

Hokkaido University, Sapporo Japan mizuta@iic.hokudai.ac.jp

Abstracts

In this article, we discuss methods for analysis of distribution valued data as a kind of symbolic data. Most methods for data analysis assume that the data are sets of numbers with structure. For example, typical multivariate data are identified as a set of n vectors of real numbers and dissimilarity data on pairs of n objects are as matrix. However, requests for analysis of data with new models become higher, as the kind and quantity of the data is increased. In order to overcome this problem, Symbolic Data Analysis (SDA) supplies various data descriptions; interval valued data, modal interval data, categorical data, distribution valued data, etc. Distribution valued data is fruitful because of its ability of expression. Professor Diday quoted Schweizer as saying "distributions are the number of the future". There are mainly two approaches to analyze distribution valued data; the use of parameters of a specific family of distributions and the use of functions (density function, cumulative distribution function, quantile function etc.) We focus on the latter and adopt quantile functions to describe the objects or data. Another great approach to deal with such complex data is Functional Data Analysis (FDA.) In FDA, the objects or data are described by functions. Then we can use techniques of FDA to analyze the quantile functions. We propose methods of clustering and multidimensional scaling (MDS) for distribution valued data with FDA techniques.

Key Words: Symbolic data analysis, functional data analysis, quantile function